DOCUMENT RESUME

ED 185 636 CS 502 916

TITLE Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, January 1 - March 31, 1980.
INSTITUTION Haskins Labs., New Haven, Conn.
PUB DATE 80
NOTE 224p.

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS Acoustics; *Articulation (Speech); Beginning Reading; Communication (Thought Transfer); Consonants; Discourse Analysis; Hearing Impairments; Language Acquisition; Oral Language; Perception; Perceptual Motor Learning; Phonetics; *Speech Communication; *Speech Skills; Vowels; Word Recognition
IDENTIFIERS *Communication Research

ABSTRACT

This report is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Drawn from the period of January 1 to March 31, 1980, the 14 manuscripts cover the following topics: (1) acoustics in human communication; (2) motor-sensory feedback formulations; (3) phonetic representation and speech synthesis by rule; (4) relationships between speech perception and speech production in normal hearing and hearing impaired subjects; (5) accessibility of the voicing distinction for learning phonological rules; (6) the effect of vocalic context on fricative perception; (7) a vibratory systems analysis of human movement production; (8) perception of limb position; (9) perceiving phonetic segments; (10) reading, linguistic awareness, and language acquisition; (11) a range-frequency effect on perception of silence in speech; (12) perception of stop consonants before low unrounded vowels; (13) apractic syndromes; and (14) a comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185636

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SR-61 (1980)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 January - 31 March 1980

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

CS502916

CONTENTS

I. Manuscripts and Extended Reports

Acoustics in Human Communication: Evolving Ideas About the Nature of Speech--Franklin S. Cooper 1

Motor-Sensory Feedback Formulations: Are We Asking the Right Questions?--J. A. Scott Kelso 9

Phonetic Representation and Speech Synthesis by Rule--Ignatius G. Mattingly 15

Relationships Between Speech Perception and Speech Production in Normal Hearing and Hearing-Impaired Subjects--Katherine S. Harris and Nancy S. McGarr 23

Accessibility of the Voicing Distinction for Learning Phonological Rules--Alice F. Healy and Andrea G. Levitt 47

Influence of Vocalic Context on Perception of the [ʃ]-[s] Distinction: II. Spectral Factors--Bruno H. Repp and Virginia A. Mann 65

Exploring a Vibratory Systems Analysis of Human Movement Production--J. A. Scott Kelso and Kenneth G. Holt 85

Properties of Slowly Adapting Joint Receptors Do Not Readily Predict Perception of Limb Position--Wynne A. Lee and J. A. Scott Kelso 109

Perceiving Phonetic Segments--Michael Studdert-Kennedy 123

Reading, Linguistic Awareness and Language Acquisition--Ignatius G. Mattingly 135

A Range-Frequency Effect on Perception of Silence in Speech--Bruno H. Repp 151

Perception of Stop Consonants Before Low Unrounded Vowels--Ignatius G. Mattingly and Andrea Levitt 167

Toward a Theory of Apractic Syndromes--J. A. Scott Kelso and Betty Tuller 175

Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences--Steven B. Davis and Paul Mermelstein 195

II. Publications 223

III. Appendix: DTIC and ERIC numbers (SR-21/22 - SR-59/60) 225

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994

National Institute of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant BNS76-82023
Grant MCS79-16177
Grant BNS78-27331

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617

National Institute of Arthritis, Metabolism, and Digestive Diseases
Grant AM25814

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Thomas Baer
Fredericka Bell-Berti+
Catherine Best+
Gloria J. Borden*
Susan Brady*
Guy Carden*
Robert Crowder*
William Ewan*
Carol A. Fowler*
Jane H. Gaitenby
Katherine S. Harris*
Alice Healy*
Leonard Katz*
Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Anders Löfqvist[2]
Virginia Mann+
Charles Marshall
Ignatius G. Mattingly*
Nancy S. McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Robert Verbrugge*
Hirohide Yoshioka[1]

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Vincent Gulisano
Donald Hailey
Terry Halwes
Sabina D. Koroluk
Agnes M. McKeon
Nancy O'Brien
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
David Zeichner

Students*

David Dechovitz
Laurie Feldman
Hollis Fitch
Carole E. Gelfer
David Goodman
Janette Henderson
Charles Hoequist
Kenneth Holt
Robert Katz
Aleksandar Kostic
Peter Kugler
Anthony Levas
Harriet Magen
Roland Mandler
Suzi Pollack
Patti Jo Price
Sandra Prindle
Brad Rakerd
Arnold Shapiro
Louis G. Tassinary
Janet Titchener
Emily Tobey-Cullen
Betty Tuller
N. S. Viswanath
Douglas Whalen
Deborah Wilkenfeld

---

*Part-time
[1]Visiting from University of Tokyo, Japan
[2]Visiting from Lund University, Sweden
+NIH Research Fellow

I. MANUSCRIPTS AND EXTENDED REPORTS

ACOUSTICS IN HUMAN COMMUNICATION: EVOLVING IDEAS ABOUT THE NATURE OF SPEECH*

Franklin S. Cooper

When one thinks about the role of acoustics in human communication, the first thing that comes to mind is speech--and speech is all that this talk will consider. There are other roles, notably in music, architecture, and applications of speech acoustics to communications technology. These will be dealt with in a following talk by Manfred Schroeder.

Even speech is a large subject. In tomorrow's session on Fifty Years of Research in Speech Communication, the seven speakers with three hours at their disposal may be able to summarize the subject in a fairly adequate way. My assignment, for these twenty minutes, must be something else. What I hope to do is not so much to present new information as to explain for those of you who are not in speech research why those of use who are in the field find it important and interesting, even exciting.

Why important? Primarily because people like to talk, and because speech is an efficient way to communicate. The latter claim might seem surprising, since we all take speech for granted, and so tend to belittle it. An experimental test, reported by Chapanis[1] only a few years ago, showed how the time required for a cooperative task depended on the mode of communication that was used, or the combination of modes. His finding was that those modes that included speech were at least twice as fast as any of those without speech. Some ways of interpreting the same data gave even higher ratios, up to ten to one in favor of speech.

Many things about speech are interesting but, since we must choose, let us concentrate on ideas about speech: how concepts about the inner nature of this odd phenomenon have changed over the past fifty years. First, though, try to recapture the feel of that long ago time. This Society was born near the end of a decade of explosive growth in the technology of communication, due largely to the vacuum tube amplifier. Remember vacuum tubes? Things we now take for granted were brand new in the twenties: radio broadcasting, talking movies, radiotelephony across the Atlantic, the rebirth of the phonograph, even experimental--very experimental--television. Truly, the ability to amplify signals, even those as weak as speech, had consequences. Concepts have consequences, too, as we shall see.

---

*Presented in Plenary Session 2 of the 50th Anniversary Celebration of the Acoustical Society of America, at Cambridge, Massachusetts, on Tuesday, 12 June 1979.

Acknowledgment: Support for the preparation of this paper was provided by Grant HD-01994 from the National Institute of Child Health and Human Development to Haskins Laboratories and by BRS Grant RR05596.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

EARLY VIEWS OF SPEECH

A view that was commonly held in the 1920's was that speech is a kind of "acoustic stuff", annoyingly complex as to detail but essentially homogeneous on average: We might refer to this, by lunar analogy, as the "green cheese theory". Dr. Crandall[2] of the Bell Telephone Laboratories wrote, in 1917, about speech

> "as a continuous flow of distributed energy, analogous to total radiation from an optical source. This idea of speech is a convenient approximation, useful in the study of speech reproduction by mechanical means."

And it was useful. Basic problems for the telephone engineer were to find out how much of the acoustic stuff the telephone must provide in order to satisfy the listener. What range of frequencies would just suffice? What range of intensities? What signal-to-noise ratios?--and so on.

Yet this way of looking at speech had flaws. Even Crandall found it "interesting" that the vowel sounds have most of the energy, whereas the consonants carry most of the information.

However, ideas about speech evolve as new tools become available. By the late twenties, a new high-frequency oscillograph led Dr. Fletcher[3] to devote some twenty pages of his book on speech and hearing to the waveforms of various words. By the early thirties, emphasis had shifted from waveform to spectrum, with some attention to the internal structure of the sound stuff. For example, in a 1930 paper, Collard[4] put it this way:

> "...a speech sound consists of a large number of components... and it is by noting at what frequencies these prominent components occur that the brain is able to distinguish one sound from another... they are called the characteristic bands... Some sounds have only one characteristic band while others have as many as five."

A different way of thinking about speech was proposed by Homer Dudley[5] at the end of the thirties. In a classic paper on "The Carrier Nature of Speech" he explained speech to his engineering colleagues by drawing an analogy with radio waves, which are not themselves the message, but only its carrier. So with speech: The message is the sub-audible articulatory gestures that are made by the speaker; the sound stuff is only an acoustic carrier modulated by those gestures.

These ideas, novel except to phoneticians, were embodied in a communications device called the vocoder, which first analyzed the incoming speech and then recreated it at the distant terminal. The vocoder was modeled after human speech, with either a buzz (like the voice) or a hissy sound as the acoustic carrier. The gestures that comprised the message were represented by a dynamically changing spectrum--a necessary engineering compromise, but one that tended to obscure Dudley's main point about the gestural nature of speech and to re-emphasize the acoustic spectrum.

SPECTROGRAMS AND THEIR CONSEQUENCES

Indeed, the view of speech as a dynamically changing spectrum had been growing in popularity even before the vocoder was invented. Although sound spectrograms were not to have their full effect on speech research until the latter half of the forties, Steinberg[6] had published in 1934 what is, in retrospect, the first spectrogram. It showed how energy is distributed in frequency and time for the sentence, "Joe took father's shoebench out". Since this one spectrogram required several hundred hours of hand measurement and computation, we can understand why this way of representing speech remained a curiosity for so long. In fact, it was not until 1946 that sound spectrograms --and a machine that could make them in minutes--emerged from the war-time research of Bell Laboratories. But spectrograms had a profound effect on speech research. They provided, literally, a new way to look at speech, as well as new ways to think about it. One way, of course, was the familiar description in spectral terms, but with a new richness of detail[7]. Now one could hope to be precise about those "characteristic bands" that distinguish the consonants.

A second way of thinking about speech was to view the spectrogram as a road map to the articulation: Thus, the formant bars on the spectrogram told one how the vocal cavities had changed in size and shape. Gunnar Fant[8] and Stevens and House[9] did much to clarify and quantify these relationships for us.

A third way of thinking about speech was to view spectrograms simply as patterns. The richness of detail was now just a nuisance, since it obscured the underlying, simpler pattern[10].

The dynamic character of speech, so evident in spectrograms, led to the development of a research instrument called the Pattern Playback. With it, spectrograms could be turned back into sound in much the way that a player piano turns a perforated musical score back into music. My colleagues Pierre Delattre and Alvin Liberman used the Playback in a long and fruitful search for what they called the "acoustic cues" for speech perception, that is, a search for those crucial parts of the spectral pattern that told the ear what sounds had been spoken[11].

ARTICULATORY NATURE OF THE ACOUSTIC CUES

One might have expected the search for the cues to have a simple, happy ending: namely, the finding of one-to-one correspondences between unvarying parts of the pattern and the minimal units of the speech. I should note that linguists were not in agreement about how to characterize the minimal units of speech--whether as phonemes, (that is, short successive segments) or as distinctive features (that is, as co-occurring attributes of longer duration). The early work sought to relate acoustic cues to phonemes. Its outcome raised issues that are unresolved to this day, and set research on two seemingly divergent paths. Let us follow one of them to the present, then return to the 1950's to pick up the other.

The acoustic cues, when they were found, proved to be neither unvarying nor simple. For a given phoneme, the cues would change, sometimes markedly,

whenever the neighboring phonemes were changed. Further, the ways in which a phoneme changed were not readily rationalized in acoustic terms, though they made good sense in terms of the articulatory gestures. These findings, and much else, led to a production-oriented view of the nature of speech. The main points were, first, that the speaker's underlying phonemic message emerges as a complexly encoded sound stream because of the several conversions it must undergo in the process of being articulated. This being so, perception of the speech by a listener necessarily involves a decoding operation, and probably a special speech device for that purpose--a built-in option available only on the homo sapiens. But what kind of mechanism, or special decoder, might that be? One possibility is a neural linkage between the auditory analyzer of the incoming speech and the motor controller of articulation, and thence upstream to the message in linguistic form--in short, perception achieved by reference to production.

This view of speech as an encoding operation[12] has had consequences for both experimental and conceptual aspects of speech research. On the experimental side, it motivated studies of how the articulators move when one is talking, of what the muscles do to make them move and, of course, how the sounds change with articulation.

On the conceptual side, interest has focused on how gestures relate back to linguistic units. The simplest relations--for example, correspondence between a particular phoneme and the contraction of a particular muscle, or between a phoneme and a target shape of the vocal tract--these simplest relations were found to be too simple to account for the data. They share that fate in varying degree with other, less simple, relationships proposed as alternatives. Indeed, the nature of the relationship is a central question in speech research today: How is the motor control of speech organized? How do linguistic units give shape to gestures?

PERCEPTION BY AUDITORY ANALYSIS

We must now go back to the 1950's, having traced the view that speech is articulatory in its very nature. There is an alternate view that stresses the role of the listener. It asks: Are there not, in the acoustic signal and its spectrogram, objective entities that correspond to the speech sounds that one hears so clearly? If this does not hold for the relationship of cues to phonemes, might it hold for distinctive features instead? The answer, despite persistent effort, has turned out to be that it is no easier to find invariant relations between features and acoustic spectrum than it is between phonemes and spectrum. I should say here that there are respected colleagues who do not share this assessment and who feel that, since the ear must do an initial analysis of speech, it is more than reasonable to suppose that the auditory system carries that analysis all the way to the linguistic units. The question, in their view, is not whether the ear does that analysis, but only how it does it.

One line of thinking has been that invariant relationships might be found in the signal after it has been transformed by the ear. A related approach has combined articulatory and auditory considerations by looking for quantal states for which variability of the gesture has only trivial effect on the sound. These stable sounds can then serve as auditory cues[13]. The principal

mechanisms proposed for interpreting auditory cues is a set of property detectors tuned to quite a variety of acoustic aspects of the signal. Most recently, interest has focused on the neural codings and recodings which the speech signal undergoes on its way up the auditory pathways[14]. This is exciting research, and there are those who hope that tracking speech to its engram in the cortex will clarify the relationship between acoustic units and linguistic units. That is surely the central point, if we are to understand speech perception.

THE NEED FOR A MODEL

So, in tracing ideas about the nature of speech from the 1950's through the 1970's, we have found unresolved questions about the choice of minimal units, and also about whether speech "belongs", in some important sense, to the mouth or to the ear. In one sense, of course, it belongs equally to each of them since the same waveform is both output and input. For a speaker, it is both at the same time. But what are the mechanisms?

We are concerned at two levels. We need, of course, to learn about the physiological mechanisms of production and perception, and we are making good progress. We need also to understand these processes at an underlying functional level--at the level of meaningful models. Do we need separate models for production and perception? It could be that each process has its own way of relating linguistic message and speech signal. In that case, we do have two models and we explain the ambivalent nature of the acoustic cues as the compromise made long ago by mouth and ear in arriving at a set of signals for spoken language. But parsimony, and a substantial body of data, argue persuasively for one model instead of two, that is to say, for a close functional linkage between production and perception that will explain how both relate to the message that speech conveys.

MESSAGES AND COMPUTERS

What is that message? What is the nature of the information contained in speech? This question is by no means new, but for some of us in speech research, it has acquired a new meaning as our field has begun to reach out from "laboratory speech", i.e., nonsense syllables, words, and the simplest of sentences, to everyday fluent speech. The question has emerged in sharpest form in work on speech understanding by computers[15].

This would justify a digression, if only time permitted, about human communication with machines. We are, I believe, on the leading edge of a new wave of technology as computers learn how to use spoken language. Never mind that the technology is still complex, expensive, and severely limited in what it can do. Remember the telephone: It was invented forty years too late, when telegraphy was already in use around the world. But how many of you have telegraphs in your homes and offices today?

A proper account would have to trace the early, and generally successful, efforts to synthesize speech automatically, and the parallel efforts, largely frustrated, on machine recognition; also, the related work on analysis-synthesis telephony and on waveform coding for better and cheaper communications between humans by means of machines.

In the area of ideas about speech, two things emerge from the research on how to converse with computers. The first is that fluent speech is the real problem, and a different and harder one than dealing with careful, word-by-word "laboratory speech". Indeed, the direct phonetic analysis of fluent speech may not even be possible. The second thing is that computers can understand such messages, under suitably constrained conditions, when the partial information available from phonetic analysis is supplemented by syntactic, semantic, and pragmatic knowledge.

THE NATURE OF FLUENT SPEECH

What is so different about fluent speech? Essentially, it is that the strategy is different, and that the acoustic signal is both less than, and more than, it was for laboratory speech: less, through depletion of phonetic detail; more, by accretion of acoustic cues direct to the syntax, the semantics, and the pragmatics. This can hardly be called an impoverished signal, but it is a different kind of signal, requiring new decoding techniques and new research on non-phonetic types of cues. A case in point is the current interest in prosody, where some of these cues are to be found.

This view of fluent speech gives new status to the acoustic signal, for now it has become the carrier, not merely of phonetic elements, but of language entities at all levels. This seems a heavy load--all of language--for so frail a carrier. Perhaps we should re-examine a long-held assumption about the nature of spoken messages: namely, that the speaker's message, all of it, is carefully packaged for aerial transport, then carried through the air to the ear and brain of a listener, where it is unpacked. But must all of it go through the air, or only those parts that are not already present in both heads, as information theory might suggest? To put it a little differently, speech could still perform its function if it carried no more than the recipe for making a message, just as a dandelion seed carries only the genetic code from one flowering to the next.

SUMMARY

We have seen that the principal role of acoustics in human communication is to let us talk with each other and, eventually, with our computers. For face-to-face communication, acoustics is quite adequate, but there is a large and growing technology in which acoustics, per se, plays a rather minor role. Even there, however, the way we conceptualize speech has important consequences.

We have come a long way in understanding the nature of spoken language. We have left behind the idea of speech as merely sound stuff, or even spectrum stuff that has certain characteristic frequencies. Spectrograms revealed the dynamic character of speech and hinted strongly at a dual nature: that speech is both a signal that is shaped by production and one that is tailored for perception. We have come to see speech as a carrier, at first mainly of phonetic messages, then as a carrier of cues to all levels of languages. We do not yet know how much, or how little, of the total message must actually be transported in fluent speech.

Finally, there are abiding questions that motivate much of the current research on speech:

1) What are the units?--a persistent question at all levels of language, though we mentioned only phonemes and distinctive features as minimal units.

2) What is the mechanism? Is speech truly anchored to production? or to auditory perception? Or, will this turn out to be a non-question, when we have finally arrived at a model that relates production and perception to each other, and to the message?

3) What is the message? How is it carried by speech? Indeed, is the total message carried at all, from head to head? or is it created anew in the head of the listener from a recipe provided by the speech signal?

I hope you now see why some of us find research on speech so challenging, and I ask your indulgence for my biases, some of which must be showing.

REFERENCES AND NOTES

1. A. Chapanis, "Interactive Human Communication", Sci. Amer. 232, 36-42 (1975).
2. I. B. Crandall, "The Composition of Speech", Phys. Rev. 10, Series 2, 74-76 (1917).
3. H. Fletcher, Speech and Hearing (New York, van Nostrand, 1929).
4. J. Collard, "Calculation of the Articulation of a Telephone Circuit from the Circuit Constants", Electr. Commun. 8, 141-163 (1930).
5. H. Dudley, "The Carrier Nature of Speech", Bell System Tech. J. 19, 495-515 (1940).
6. J. C. Steinberg, "Application of Sound Measuring Instruments to the Study of Phonetic Sounds", J. Acoust. Soc. Am. 6, 16-24 (1934).
7. R. K. Potter, G. A. Kopp and H. C. Green, Visible Speech (New York, van Nostrand, 1947).
8. C. G. M. Fant, Acoustic Theory of Speech Production (The Hague, Mouton, 1960).
9. K. N. Stevens and A. S. House, "Development of a Quantitative Description of Vowel Articulation", J. Acoust. Soc. Am. 27, 484-493 (1955); ibid., "Studies of Formant Transitions Using a Vocal Tract Analog", J. Acoust. Soc. Am. 28, 578-585 (1956).
10. F. S. Cooper, A. M. Liberman and J. M. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech", Proc. Nat. Acad. Sci. 37, 318-328 (1951).
11. An informal account of these early experiments and their interpretation is given by A. M. Liberman and F. S. Cooper, "In Search of the Acoustic Cues", in Melange a la Memoire de Pierre Delattre, edited by A. Valdman (The Hague: Mouton, 1972), 9-26.
12. For a review of the experimental data and a mid-course interpretation, see A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code", Psych. Rev. 74, 431-461 (1967); for an updated interpretation, A. M. Liberman and M. Studdert-Kennedy, "Phonetic Perception", in Handbook of Sensory Physiology, Vol. VIII:

Perception, edited by R. Held, H. Leibowitz and H-L. Teuber (Heidelberg, Springer Verlag, 1978), 143-178.

13. K. N. Stevens, "The Quantal Nature of Speech: Evidence from Articulatory-acoustic Data", in Human Communication, a Unified View, edited by P. B. Denes and E. E. David (New York, McGraw-Hill, 1972), 51-66; S. E. Blumstein and K. N. Stevens, "Acoustic Invariance in Speech Production: Evidence from Measurements of the Spectral Characteristics of Stop Consonants", J. Acoust. Soc. Am., 66, 1001-1017 (1979); K. N. Stevens and S. E. Blumstein, "The Search for Invariant Acoustic Correlates of Phonetic Features", in Perspectives on the Study of Speech, edited by P. D. Eimas and J. Miller (New Jersey, Erlbaum Assoc.) (In press).
14. These ideas are discussed at this 50th Anniversary Meeting in the invited papers of Session NN by N. Y-S. Kiang, K. N. Stevens, B. Delgutte, and M. B. Sachs and E. D. Young.
15. D. H. Klatt, "Review of the ARPA Speech Understanding Project", J. Acoust. Soc. Am. 62, 1345-1366 (1977).

MOTOR-SENSORY FEEDBACK FORMULATIONS: ARE WE ASKING THE RIGHT QUESTIONS?*

J. A. Scott Kelso+

Gyr and his colleagues (1979) would have us confront anew the evidence for assigning a critical role for motor activity in visual perception. While their discussion of "sensorimotor processes" suggests a tightly coupled relationship between perception and action systems, we believe that the authors present a potentially misleading picture of the relationship between efference and afference. The thrust of this article will be to lay out some of the logical problems associated with a theory that utilizes the concept of efference copy. By and large these are in addition to those already raised by a number of commentators on a paper by Roland (1978). The present position is that efference copy and its often synonymously used affiliates, corollary discharge and central monitoring of efference are, with perhaps a single exception, low on theoretical power. This is the general claim to be made here. In addition, we wish to point specifically to an alternative account for the type of data that Gyr et al. seek to explain. We focus on the concepts of information discordance and allocation of attention, which when allied, seem to provide an adequate explanation of much of the adaptation literature--without resorting to unique contributions from efference copy.

Gyr et al. present the classical data for the role of efference in visual perception. Many of the methodological problems in extending this approach to human behavior have been raised by Shebilske (1977) and will not be re-enumerated here. More damaging, however, is the elegant rationale by Turvey (1977a) that an explanation of visual perception relying on a comparison of efferent signals to eye muscles and the retinal input provided by vision, falls sadly short when we move beyond the situation of a simple eye movement in a stationary head on a stationary body. When one considers the complexity of the visual array when an individual performs locomotory activities, for example, a simple computational explanation no longer suffices.

It is clear that Gyr et al. wish to extend the efference copy notion to movement coordination in general. They are, of course, not alone in this enterprise in that the efference copy concept is often used to explain--among other things--(1) the superiority of active over passive movement perception (e.g., Kelso, 1977; Teuber, 1974); (2) subject's ability to make rapid error corrections in step tracking tasks well within the bounds of peripheral

---

*This paper was an invited commentary on Gyr, J., Willey, R., and Henry, A. Motor-sensory feedback and geometry of visual space. The Behavioral and Brain Sciences, 1979, 2, 59-74.

+Also University of Connecticut.

Acknowledgment: I thank Kenneth Holt and James Pruitt for their assistance in the preparation of this paper, which was supported by NIH Grants AM 25814 and NS 13617.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

feedback loop times (e.g., Higgins & Angel, 1970); and (3) the motor performance of deafferented animals (e.g., Taub, 1977). While such data require satisfactory explanation, we do not want to place our money on an all-encompassing efference copy-reafference relationship. As long ago pointed out by Bernstein (1967), there is an equivocality between motor commands and the effects they produce. Therefore, there can be no direct comparison between efference copy and reafference because such a one-to-one mapping between the two sources of information cannot exist.

More important for a theory of coordination is the issue of how the multiple degrees of freedom of the motor apparatus are regulated. Powerful arguments can be generated against a view that efferent commands specify the states of individual muscles. A consequence of such a view would be an extraordinarily detailed efference copy that fails to take advantage of the intrinsic organization of the nervous system (for details see Grillner, 1975). Rather we wish to view efference not in an executive role but as organizational, in which the entities regulated are coordinative structures (Easton, 1972; Turvey, 1977b); that is, functional groupings of muscles that are constrained to act as a single unit.

A specific operation of efference in this perspective is feedforward in nature such that the performer is prepared for the impending motor output and the afference arising from such activity. Thus various experiments have illustrated postural adjustments and descending biasing influences on the segmental machinery in preparation for particular types of activity such as lifting the arm or dorsiflexing the foot (see Kots, 1977). Note that efference does not necessarily carry a central, motor-to-sensory corollary discharge connotation (Teuber, 1974). Such a view, while placing the motor commands in a sensory 'code' readily available for comparison with reafference, is just as subject to the mapping invariance and degrees of freedom criticisms outlined above. Rather, efference may be viewed in terms of feedforward, which, because of its particular biasing or tuning operations on the spinal cord, constrains the performer to a limited set of activities (Fowler, 1977; Greene, 1972).

Gyr et al. resort to deafferentation research as evidence for autoregulation of behavior at a central level. In agreement with Pew (1974) we would have to say that the argument is a default one taking the following form: (1) peripheral feedback has been eliminated, (2) the animal can perform various motor activities, (3) therefore some internal monitoring mechanism is responsible. A variety of alternative conclusions have been offered (e.g., Adams, 1976; Schmidt, 1975). But it has never been clear in this formulation what is meant by monitoring or the nature of the entity that is being monitored. Taub's more recent work on perinatal deafferentation (e.g., Taub, 1977 for review) can be interpreted to mean that residues of past experiences, efference copies and the like, are unsuitable candidates for the monitored representation. These are likely to be very impoverished indeed and hardly able, even if one could imagine them to do so, to contain all the details of the action patterns such as climbing, hanging and grasping, that have been observed. But the stronger criticism here is that posing the question: Is an efferent signal necessary or not for normal perception? is a conceptual error. The tight coupling between efference and afference demands that we not treat them as individual entities but rather seek to understand the nature of

their interaction. Some headway has already been made in this regard. There is neurophysiological evidence that prior to and during voluntary movements in cats, afferent information in the dorsal column medial lemniscus is modified (Ghez & Lenzi, 1971; Coulter, 1974). Similarly, anatomical evidence reveals that descending pyramidal fibers exert both pre- and post-synaptic influences on the transmission of sensory information in the spinal cord (Kostyuk & Vasilenko, 1968). Furthermore, human psychophysical experiments on the perception of vibratory stimuli show that the sensory threshold becomes elevated during voluntary movement (Dyhre-Poulson, 1975). This modulation is specific to the digit being moved and is not merely a general gating effect on sensory inputs. In sum, we have evidence from a variety of sources illustrating the efferent modulation of afference.

Just as interesting is the rather direct influence of afferent information on efferent activity. At a neurophysiological level, Easton (1972) has shown that stretch of the vertical eye muscles leads to facilitation and inhibition of cat forelimb flexor and extensor muscles. A downward directed gaze resulted in facilitated forelimb extension while upward gaze facilitated flexion. More recently, Thoden, Dichgans, and Savidis (1977) have produced evidence that hindlimb flexor and extensor activity can be modulated by both vestibular and visual stimulation. Of particular note is the finding that direction-specific reflex excitability in extensor and flexor motoneurons could be induced by rotating a visual display about the cat's line of sight. Thus counterclockwise rotation, indicating displacement to the right, led to an enhancement of extensor motoneuronal activity and a depression in flexor motoneurons, while clockwise rotation had an equal but opposite effect. Analogous findings are available from the elegant "swinging room" experiments of Lee and his colleagues (see Lee, 1978). Even though the subject is supplied with veridical information from kinesthetic receptors that the floor is stable, posture and balance are shown to be under visual control as evident in the excessive sway observed when the room is moved. Indeed body sway can be visually driven by oscillations as small as 6 mm without the subject's being aware of it. All this points to a tight coupling, a specification as it were, of efference by afference.

The general claim here, then, is that the efference copy construct cannot handle the vagaries of the motor system nor does it provide a particularly useful explanatory device for visual perception. Neither do we want to approach the issue of adaptation via a framework that promotes a dichotomy between efference and afference as Gyr et al. have done. In actuality, there is no need to revert to a recorrelation formulation for an explanation of perceptual adaptation. It is now well-documented, for example, that adaptation can occur without movement (Howard, Craske, & Templeton, 1965), in passive conditions (Melamed, Halay, & Gildow, 1973), and in conditions where passive movement is induced by vibration (Mather & Lackner, 1975). All that is needed for adaptation to occur is a discordance between two or more sources of information that are normally congruent with each other. The performer's attempt to nullify this discordance, and hence return the inputs to their previous correspondence, is seen to be representative of the adaptive process. Numerous studies support this viewpoint (see Kornheiser, 1976, for a review) by showing that the degree to which adaptation takes place is a function of the information available to the subject regarding the altered state of the system.

While the notion of discordance is plausible as an account for the occurrence of adaptive change, it lacks predictive power with regard to the exact form that such change will take. The additional concept of attentional allocation provides a potential solution to this problem in that the outcome of any noncorrespondence between two sources of information (say proprioceptive information detected visually and proprioceptive information detected by joint, muscle and tendon receptors) can be predicted on the basis of the attentional demands of each input. Thus Canon (1970), Kelso, Cook, Olson, and Epstein (1975) and more recently, Warren and Schmitt (1978), have all shown that adaptation takes place in the modality that is not used during the exposure period. When allocation of attention is left uncontrolled, the dominant modality (in most cases, vision) will remain stable while the paired source of information will undergo an adaptive shift.

We are left then to explain, within this formulation, the consistent finding that self-produced movement facilitates the adaptive process more than passive movement. Viewed from the informational account, we would argue that under active conditions the subject is sensitized to pay attention to the discordance between the seen and felt positions of the limb, while under passive conditions, attention is more evenly distributed between the two sources of information. Given the dominance of vision and the subjects' inherent bias to attend to it (Posner, Nissen, & Klein, 1976), we would then expect greater adaptation under self-produced movement conditions. What matters then for the adaptive process is information about discordance, which, when combined with attentional factors, seems adequate to explain the findings attributed to motor-sensory mechanisms.

In the present view, therefore, there is no urgent need to reopen this issue based on Gyr et al.'s failure to replicate Held and Rekosh. Many of Held's predictions have been tested over and over again in an area already burgeoning with empirical data (e.g., Kornheiser, 1976; Welch, 1974, for reviews). The real need, therefore, is not for more experimentation but rather for more understanding of the nature of the adaptive process, with particular reference to the interaction of efference and afference.

REFERENCES

Adams, J. A. Issues for a closed-loop theory of motor learning. In G. E. Stelmach (Ed.), Motor control: Issues and trends. New York: Academic Press, 1976, 87-108.

Bernstein, N. The coordination and regulation of movement. New York: Pergamon, 1967.

Canon, L. K. Intermodality inconsistency of inputs and directed attention as determinants of the nature of adaptation. Journal of Experimental Psychology, 1970, 84, 141-147.

Coulter, J. D. Sensory transmission through leminiscal pathway during voluntary movement in the cat. Journal of Neurophysiology, 1974, 37, 831-845.

Dyhre-Poulsen, P. Increased vibration threshold before movements in human subjects. Experimental Neurology, 1975, 47, 516-522.

Easton, T. On the normal use of reflexes. American Scientist, 1972, 60, 591-599.

Fowler, C. A. Timing control in speech production. Bloomington, Ind.: Indiana University Linguistics Club, 1977.

Ghez, C., & Lenzi, G. A. Modulation of sensory transmission in cat lemniscal system during voluntary movement. European Journal of Physiology, 1971, 323, 272-278.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), Progress in theoretical biology (Vol. 2). New York: Academic Press, 1972.

Grillner, S. Locomotion in vertebrates: Central mechanisms and reflex interaction. Physiological Reviews, 1975, 55, 247-304.

Gyr, J., Willey, R., & Henry, A. Motor-sensory feedback and geometry of visual space. The Behavioral and Brain Sciences, 1979, 2, 59-94.

Higgins, J. R., & Angel, R. W. Correction of tracking errors without sensory feedback. Journal of Experimental Psychology, 1970, 84, 412-416.

Howard, I. P., Craske, B., & Templeton, W. B. Visuomotor adaptation to discordant exafferent stimulation. Journal of Experimental Psychology, 1965, 70, 189-191.

Kelso, J. A. S. Planning and efferent components in the coding of movement. Journal of Motor Behavior, 1977, 9, 33-47.

Kelso, J. A. S., Cook, E., Olson, M. E., & Epstein, W. Allocation of attention and the locus of adaptation to displaced vision. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 237-245.

Kornheiser, A. S. Adaptation to laterally displaced vision: A review. Psychological Bulletin, 1976, 83, 783-815.

Kostyuk, P. G., & Vasilenko, D. A. Transformation of cortical motor signals in spinal cord. Proceedings of the IEEE, 1968, 56, 1049-1058.

Kots, Ya. M. The organization of voluntary movement. New York: Plenum, 1977.

Lee, D. L. The functions of vision. In H. L. Pick & E. Saltzman (Eds.), Modes of perceiving and processing information. Hillsdale, N.J.: Erlbaum, 1978.

Mather, J., & Lackner, J. Adaptation to visual rearrangement elicited by tonic vibration reflexes. Experimental Brain Research, 1975, 24, 103-105.

Melamed, L. E., Halay, M., & Gildow, J. W. Effect of external target presence on visual adaptation with active and passive movement. Journal of Experimental Psychology, 1973, 98, 125-130.

Pew, R. W. Human perceptual-motor performance. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. New York: Lawrence Erlbaum Associates, 1974.

Posner, M. I., Nissen, M. J., & Klein, R. M. Visual dominance: An information-processing account of its origins and significance. Psychological Review, 1976, 83, 157-171.

Roland, P. E. Sensory feedback to cerebral cortex during voluntary movements in man. The Behavioral and Brain Sciences, 1978, 1, 129-147.

Schmidt, R. A. A schema theory of discrete motor skill learning. Psychological Review, 1975, 82, 225-269.

Shebilske, W. L. Visuomotor coordination in visual direction and position constancies. In W. Epstein (Ed.), Stability and constancy in visual perception: Mechanisms and processes. New York: Wiley, 1977.

Taub, E. Movement in nonhuman primates deprived of somatosensory feedback. In J. Keogh (Ed.), Exercise and sports sciences reviews (Vol. 4). Santa Barbara: Journal Publishing Affiliates, 1977, 335-374.

Teuber, H-L. Key problems in the programming of movements. Brain Research,

1974, 71, 533-568.

Thoden, V., Dichgans, J., & Savidis, T. Direction-specific optokinetic modulation of monosynaptic hind limb reflexes in cats. Experimental Brain Research, 1977, 30, 155-160.

Turvey, M. T. Contrasting orientations to the theory of visual information processing. Psychological Review, 1977, 84, 67-78. (a)

Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing. Hillsdale, N.J.: Erlbaum, 1977, 211-266. (b)

Warren, D. H., & Schmitt, T. L. On the plasticity of visual-proprioceptive bias effects. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 302-310.

Welch, R. B. Research on adaptation of rearranged vision: 1966-1974. Perception, 1974, 3, 367-392.

PHONETIC REPRESENTATION AND SPEECH SYNTHESIS BY RULE*

Ignatius G. Mattingly+

While a computer model of performance in speech production is certainly conceivable, the title of this seminar, "Speech Synthesis Programs as Models of Speech Production," seems incorrect as a characterization of existing systems for synthesis of speech by rule. Insofar as these systems have had more than the purely pragmatic goal of converting written utterances into spoken utterances, the objective has been a more modest one, but one that is nonetheless directly related to the theme of this Symposium, "The Cognitive Representation of Speech." The objective has been to elucidate the nature of the phonetic representation shared by speaker and hearer (and recorded, after a fashion, in a conventional phonetic transcription) and the relationship of this representation to physical events. A simulation of the mental activity underlying actual speech production would be a far more ambitious project, though one well worth attempting.

Synthesis by rule has also, of course, considered other interesting questions: the relationship of the conventional orthography to the phonemic representation, and the relationship of the latter to the phonetic representation. But its central concern, surely, has been to relate the phonetic representation either to the articulatory or to the acoustic parameters of continuous speech. The practitioner of synthesis by rule cannot avoid this concern, for his rules must somehow specify the physical correlates of the elements of the phonetic representation, and his program must necessarily include some kind of algorithm for getting from one phonetic element to the next.

The simplest conceivable synthesis-by-rule system would be one in which each phonetic element is associated with an articulatory or acoustic segment of specified duration (or of a sequence of such segments). Since the inventory would be small, prerecorded segments of natural speech rather than parametrically-specified synthetic segments could be used. This system would have only one "rule": concatenate segments corresponding to successive

---

*Paper presented at the International Symposium on the Cognitive Representation of Speech, Edinburgh, July 29-August 1, 1979. To appear in T. Myers, J. Laver, and J. Anderson (Eds.), The Cognitive Representation of Speech. Amsterdam: North Holland.

+Also University of Connecticut.

Acknowledgment: Preparation and presentation of this paper was supported in part by grant HD-01994 from the National Institute of Child Health and Human Development and in part by a travel grant from the University of Connecticut Research Foundation.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

phonetic elements. Such schemes have indeed been proposed (Harris, 1953), but it was appreciated quite early by practitioners of speech synthesis by rule that the acoustic record (and a fortiori the articulatory record) contained intervals that could not reasonably be viewed as steady-state segments correlated with single phonetic elements, and that these intervals (which were regarded as "transitions" connecting "true" steady-state segments) were perceptually extremely important (Liberman, Ingemann, Lisker, Delattre, & Cooper, 1959).

Thus, the synthesis-by-rule systems developed in the 60's (see Mattingly, 1974, for a review), though differing in detail, had in common the requirement that transitions, as well as steady states, had to be described in the synthesis rules. In the synthesis-by-rule system described by Holmes, Mattingly, and Shearme (1964), for example, each "phoneme table" specifies, for each parameter, not only a steady-state value but also a contribution to a "boundary value," and the moment-by-moment values during a transition are calculated by linear interpolation from the steady-state value for the current phonetic element to the boundary value, and from the boundary value to the steady-state value for the following element. The interpretation of the phonetic representation implicit in these systems was thus consistent with the concept of segmentation proposed by Pike (1943): Speech consists of "level segments," during which one or more articulatory "crests or troughs of stricture" are maintained, connected by "glide" segments, during which one or more strictures are being applied or released. Successive elements of the phonetic representation referred to successive level segments, from which the glides could be predicted.

This approach to synthesis by rule, though demonstrably capable of producing speech that was by conventional measures intelligible (Nye & Gaitenby, 1973; Pisoni, Note 1), ran into three related sorts of difficulty. First, it can be observed in natural speech that the acoustic "steady-state" segment associated with a particular phonetic element, and likewise the transitions from the preceding and to the following segments, vary according to the immediate phonetic context in ways that are clearly of perceptual importance. For practical purposes, this variation could be dealt with by writing allophone rules to modify the phonetic specification according to context, or by allowing the transition from the preceding steady-state segment and the transition to the following steady-state segment to overlap, eliminating the current steady-state segment (Holmes et al., 1964). Either approach was an admission of a deficiency in the Pikean view of phonetic representation (from an articulatory standpoint, the deficiencies of the Pikean view are even more glaring since the simultaneous maintenance of several constrictions over a period of time--an articulatory "steady-state"--is quite unusual).

Variations that depended on non-adjacent phonetic context were less easily dealt with. Since the algorithms used in these systems considered only two consecutive phonetic elements at a time, this kind of contextual variation could be handled only by frankly ad hoc procedures. Though it was generally assumed that such variation was on the whole of little perceptual importance, its very existence posed further problems for the Pikean view.

The final, and theoretically most damning, difficulty with the Pikean approach was that the durations of steady-state and transitional segments are

subject to extensive contextual variation and had to be assigned completely ad hoc. Even these ad hoc assignments were notoriously unsuccessful in producing realistic speech-timing patterns. But what is at issue is not just whether the durations for phonetic elements could be adequately specified within a segment-by-segment framework (or even by a more elaborate framework involving higher-order prosodic units), but whether the elements of a phonetic representation can be said to have durations at all.

The fact that some synthesis-by-rule systems have respectable intelligibility scores should not lead anyone to suppose that these problems are of no practical importance. It is clear that the speech produced by these systems, though intelligible, places a much greater load on short-term memory than does natural speech. Nye and Gaitenby (1974) investigated the ability of listeners to recall, immediately after presentation, semantically anomalous but syntactically acceptable sentences (e.g., "The wrong shot led the farm") under two conditions. In one condition, listeners heard naturally-spoken sentences; in the other, sentences synthesized with the Haskins Laboratories synthesis-by-rule system. Pisoni (Note 1) later used the same test in an evaluation of Klatt's (1979) synthesis-by-rule system. The percentages of words correctly reported were: for natural speech, 95%; for the Haskins system (Mattingly, 1968; Kuhn, 1973), 78%; and for the Klatt system, 78.7%. Though both synthesis-by-rule systems have conventional intelligibility scores close to natural speech, it would seem that the unnaturalness of synthetic speech, in particular, perhaps, the failure to represent coarticulatory variation adequately and the unnaturalness of the timing of acoustic events, seriously interferes with short-term memory coding.

The objections to segmental models have often been pointed out, and it is rather surprising that practitioners of synthesis by rule have paid so little attention. Menzareth and Lacerda (1933), Joos (1948), Fant (1962), Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) and many others have made it clear that a many-to-many relationship between phonetic elements and acoustic segments, however defined, is the norm rather than the exception in speech, and that--in Fant's words--"several adjacent sounds [i.e., acoustic segments] may carry information on one and the same phoneme, and there is overlapping insofar as one and the same sound segment carries information on several adjacent phonemes" (1962:9). Nor is it the case that the many-to-many relationship is to be attributed solely to the merging into a single acoustic stream of effects due to several articulators, for multiple phonetic influences simultaneously affect the movements of an individual articulator (MacNeilage & Sholes, 1964; MacNeilage & DeClerk, 1969).

In other words, information in speech is transmitted in parallel, and it is this fact that makes possible higher information rates for speech than would be possible in a truly segmental process. But parallel transmission appears to present no difficulty for the speech-perception mechanism. On the contrary, this mechanism seems to be specialized for decomposing an encoded period of speech into the component phonetic influences, and is rather less at home with isolated consonants or isolated vowels (Liberman et al., 1967).

Both acoustic and articulatory models to account for the many-to-many relationship between phonetic elements and acoustic events have been proposed. Joos regarded phonetic elements as "overlapping innervation waves" and showed

that the formant trajectory for a C_1VC_2 syllable could be decomposed into a vowel "layer," an initial-consonant layer and a final-consonant layer (Joos, 1948:109,125). In the model proposed by Ohman (1967) to account for observed dynamic changes in vocal-tract shape in V_1CV_2 syllables, the predicted shape depends on a time function, a vowel-shape function, a consonant-shape function, and a coarticulation function associated with the consonant. Even for values of the time function where the consonant-shape function predominates, the predicted shape function depends also on the vowel function, to the extent determined by the value of the coarticulation function. Moreover, if the vowel-shape function is not constant but varies depending on the time-varying values of phonetic features, that is, if $V_1 \neq V_2$, the predicted shape at any point in the V_1CV_2 sequence depends on these feature-value time functions as well as on the consonant-shape and coarticulation functions. An observed vocal-tract shape during a VCV utterance is thus interpreted as the result of superimposition of a consonant shape tolerating a certain degree of coarticulation upon a changing vowel shape.

It is worth noting that the Joos and Ohman models are essentially "prosodic." That is, they treat elements of the so-called segmental sequence in the way that prosodic features are conventionally treated. The overlapping and layering of prosodic features is usually taken for granted: It would be quite unconventional to propose a division of the signal into stress and intonation segments. These models are also quite consistent with earlier attempts by phoneticians and phonologists to treat one or another "segmental" element as if it were a prosodic feature (see, for example, Hockett's discussion of "componential analysis," 1955:129 ff).

However, it is not sufficient to regard the sounds of speech merely as an inventory of "innervation functions" or (as we prefer to call them) "phonetic-influence functions" that may overlap with one another freely and to an indefinite extent. Phonetic elements are perceived as ordered, and if this ordering is not to be attributed to the existence of successive segments, some other explanation is required. Moreover, there are obvious restrictions on the co-occurrence of overlapping patterns, and corresponding restrictions on the perceived ordering of phonetic elements.

The basis for these restrictions becomes obvious if we consider what combinations of phonetic influences can in fact be effectively transmitted in parallel. If, in the utterance [pla], the onset, constriction and release for [l] were to occur entirely during the period of closure for [p], the [l] would have no acoustic correlates. But if the [l] release is delayed until after the [p] release is well advanced, information about [l] (as well as about [p] and [a]) is available both before and after the [l] release. There has to be some means of guaranteeing that this second pattern will in fact be the one that is used. Again, in stop sequences of the form $V_1S_1 \ldots S_nV_2$, information about stops $S_2 \ldots S_{n-1}$ will be present if the release of S_j is delayed relative to that of S_{j-1}. But the period of constriction for S_j, because the constriction is maximally close, will convey only manner information, and the burst will convey place information about S_j itself, but little or no information about any other phonetic element. Hence there will be no effective parallel transmission except for the periods when the S_1 constriction is being applied during the constriction for V_1 and the S_n constriction is being released during the constriction for V_2. Thus length of stop

sequences has to be severely limited, as is the case is all languages.

The general articulatory prerequisite for parallel transmission would appear to be that the constrictions for one or more closer articulations must be in the process of being released or applied in the presence of constrictions for one or more less close articulations. In terms of this formulation, the conventional ranking of manner classes according to degree of closeness (obstruents, nasals, liquids, glides, vowels) corresponds to a ranking according to the degree to which information can be encoded during the release or application of the constriction, and the inverse of this ordering, to the degree to which information can be encoded during the period of maximal constriction. [Holmes et al. (1964) exploited this ranking of the manner classes to a limited extent in their synthesis-by-rule system.] If parallel transmission is to be maximized, then the articulations of speech must be scheduled so that periods during which constrictions are released in rank order alternate with periods during which constrictions are applied in inverse rank order. This is of course exactly what is accomplished by the syllabic organization of speech. It would seem, therefore, that the syllable has more than a phonological or prosodic role: It is the means by which phonetic influences are scheduled so as to maximize parallel transmission.

The perception that phonetic elements are ordered thus has an obvious explanation. This perception does not arise from the detection of successive segments, or even of the successive releases or applications of constrictions. It is rather the ranking of the manner classes itself that governs the percept. That is, the listener interprets the available acoustic data in terms of a framework of expectations about the structure of the syllable based on the ranking of manner classes.

Interpreting syllable structure in terms of the manner-class ranking is in itself hardly novel: Jespersen (1926) proposed such a ranking, based on "sonority," as the basis for an account of the syllable. But the argument here is that if segments are to be replaced in a phonetic model by phonetic-influence functions, syllable structure is essential for efficient speech communication and is not simply a concomitant linguistic structure.

At Haskins Laboratories, we are developing a new synthesis-by-rule system in which acoustic parameters depend on the interaction of overlapping phonetic influences, and the timing of these influences is determined by the structure of phonetic syllables. For various practical reasons, we have chosen to use the acoustic parameters of a terminal-analog synthesizer rather than articulatory ones, but an essentially similar approach could be used with an articulatory synthesizer.

The phonetic elements that are considered to influence the acoustic character of a syllable in our system are the vowels of the current, preceding and following syllables, the initial consonants of the current and following syllables, and the final consonants of the current and preceding syllables (higher-level prosodic elements have not as yet been taken into consideration). With each such phonetic influence is associated a rank that depends upon the manner class of the element, and within manner class, upon temporal order; a set of target parameter values; and a time-function, ranging in value between 0 and 1, which represents the weight of the influence relative to the

combined weight of all lower-ranking influences.

A phonetic-influence function is defined from the beginning of the preceding syllable to the end of the current syllable, in the case of syllable-initial articulations, or from the beginning of the current syllable to the end of the following syllable, in the case of syllable-final articulations (in this way, intersyllabic influences are taken care of). An influence function has a growth period, during which it has the form $I_t = \kappa e^{\beta t}$ (cf. Lindblom, 1963), a possible steady-state period of duration h during which $I_{t=1}$, and a declining period during which $I_t = \kappa e^{-\gamma(t-h)}$. The rate at which an influence grows (or declines) depends on β (or γ), its effective onset time on κ. Syllables may vary in duration according to their phonetic structure and the value of κ is adjusted accordingly.

Given $I_{i,t}$, the strength of the ith-ranking influence at time t, and $T_{i,j}$, the target value for the jth parameter associated with this influence, the parameter value reflecting the ith and lower-ranking influences is

$$V_{i,t,j} = V_{i-1,t,j} + I_{i,t}(T_{i,j} - V_{i-1,t,j})$$

Taking as $T_{0,j}$ the target value for the vowel of the previous syllable, the parameter value reflecting all influences can be calculated iteratively. At any particular instant, the weight of most possible influences will be zero or near zero, and computation is speeded by neglecting these influences.

The variables of this algorithm that are associated with influences of elements of each manner class are defined by an ordered set of rules. These variables include the target parameter values, the increment to syllable duration attributable to the element, the duration of the steady-state period, the times relative to the notional beginning (or end) of a syllable, when the strength of an initial (or final) influence equals .5, 1, and .5 again (κ, β and γ are determined from these time-values). The definitions of these variables in the rules are conditional upon particular patterns of feature-values that might be specified in the phonetic description of the syllable. Before the parameter values are computed, the pattern of feature-values in each rule is compared with the actual phonetic description. If the rule applies, the algorithmic variables mentioned are defined according to the rules. Since the rules are ordered, a variable may well be redefined by one or more subsequent rules.

We feel that this scheme reflects more clearly the essential character of the relationship between the phonetic representation and acoustic events than our earlier synthesis-by-rule system, or other systems in which a Pikean segmentation is assumed. We hope that it will make possible the production of at least equally intelligible and more natural and more understandable synthetic speech.

REFERENCE NOTE

1. Pisoni, D. B. Some measures of intelligibility and comprehension. Chapter prepared for MIT summer course, "Conversion of Unrestricted English Text to Speech," June, 1979.

REFERENCES

Fant, G. Descriptive analysis of the acoustic aspects of speech. Logos, 1962, 5, 3-17.

Harris, C. M. A study of the building blocks of speech. Journal of the Acoustical Society of America, 1953, 25, 962-969.

Hockett, C. F. A manual of phonology (Memoir 11). International Journal of American Linguistics, 1955, 21, 1955, No. 4, Part 1.

Holmes, J. N., Mattingly, I. G., & Shearme, J. N. Speech synthesis by rule. Language and Speech, 1964, 7, 127-143.

Jesperson, O. Lehrbuch der Phonetik. Leipzig: Teubner, 1926.

Joos, M. Acoustic phonetics. (Monograph 23). Language, 1948, 24, No. 2, Suppl.

Klatt, D. Structure of a phonological rule component for a synthesis-by-rule program. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, ASSP 24, 391-398.

Kuhn, G. M. A two-pass procedure for synthesis by rule. Journal of the Acoustical Society of America, 1973, 54, 339. (Abstract)

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P. C., & Cooper, F. S. Minimal rules for synthesizing speech. Journal of the Acoustical Society of America, 1959, 31, 1490-1499.

Lindblom, B. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 1963, 35, 1773-1781.

MacNeilage, P. F., & DeClerk, J. L. On the motor control of coarticulation of CVC monosyllables. Journal of the Acoustical Society of America, 1969, 45, 1217-1233.

MacNeilage, P. F., & Sholes, G. N. An electromyographic study of the tongue during vowel production. Journal of Speech and Hearing Research, 1964, 7, 209-232.

Mattingly, I. G. Experimental methods for speech synthesis by rule. IEEE Transactions on Audio, 1968, 16, 198-202.

Mattingly, I. G. Speech synthesis for phonetic and phonological models. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 12). The Hague: Mouton, 1974.

Menzerath, P., & Lacerda, A. de. Koartikulation, Steurung und Lautabgrenzung. Berlin: F. Dümmler, 1933.

Nye, P. W., & Gaitenby, J. Consonant intelligibility in synthetic speech and in a natural speech control (Modified Rhyme Test results). Haskins Laboratories Status Report on Speech Research, 1973, SR-33, 77-91.

Nye, P. W., & Gaitenby, J. The intelligibility of synthetic monosyllable words in short, syntactically normal sentences. Haskins Laboratories Status Report on Speech Research, 1974, SR-37/38, 169-190.

Ohman, S. Numerical models of coarticulation. Journal of the Acoustical Society of America, 1967, 41, 310-320.

Pike, K. Phonetics. Ann Arbor: University of Michigan Press, 1943.

RELATIONSHIPS BETWEEN SPEECH PERCEPTION AND SPEECH PRODUCTION IN NORMAL HEARING AND HEARING-IMPAIRED SUBJECTS*

Katherine S. Harris+ and Nancy S. McGarr++

The purpose of this paper is to consider the factors that make deaf speech so unintelligible to listeners. Twenty years of research have generated a long list of aspects of deaf production that are characteristic of the group, which become increasingly deviant with standard measures of hearing loss, and which are correlated with overall intelligibility. However, while all of these aspects may make the deaf individual identifiable, they may not contribute equally to the listener's ability to understand him or her.

Traditionally, training problems have focused heavily on static articulator placement. More recently, teachers have attempted to correct some deviant suprasegmental characteristics of the speech, on the assumption that as these characteristics become more normal, the speech will improve. In this paper, we will argue, first, that although the suprasegmental characteristics of deaf speech may be grossly abnormal overall, the pattern of production gives evidence that deaf speakers "know" some of the rules for patterning fundamental frequency and duration. Indeed, the correction of those particular aspects of production that are deviant does not result in enormous improvements in intelligibility. We will argue, further, that deaf speech is deviant in ways that arise from an inability to coordinate the temporal relations among articulators. This results in a pattern of production that cannot be appropriately classed as deviant simply in its segmental or suprasegmental aspects, and, indeed, is difficult to characterize within a conventional descriptive matrix.

A MODERN VIEW OF SPEECH TRAINING

No account of speech production in the deaf can avoid beginning with the pivotal work of Daniel Ling (1976), who has attempted to systematize training based on a rationale derived from the modern speech science literature. At

---

*A preliminary version of this paper was presented at the Institute on Speech Assessment and Speech Improvement in the Hearing Impaired, National Technical Institute for the Deaf, Rochester, New York, June 22, 1979, and is to appear in Proceedings of the Institute, Washington, D.C.: Alexander Graham Bell Association, in press. The experimental results were described in a paper presented at the meeting of the Acoustical Society of America, Cambridge, Mass., June, 1979.

+Also Graduate School and University Center, the City University of New York.
++Also Molloy Catholic College for Women, Rockville Center, New York.

Acknowledgment: We are grateful to Mary Joseph Osberger and Abigail Peterson Reilly for their comments on this manuscript. The work described in this paper was supported by Grant 13617 to Haskins Laboratories.

the center of his contribution lie two great themes. The first is that speech production skills should be developed in their natural order--that is, in the order in which they develop in normal children. He suggests that breath and voice control be developed early, with the development of segmental aspects of production coming later, following the order in which sounds develop in the child.

This idea has its basis in a view of speech as overlaid on the normal functions of the articulatory mechanism. The effect of teaching the individual sounds of speech to the deaf after respiratory and phonatory control has been mastered should be that the filter function, the result of movement of the upper articulators, is superimposed on a matrix of adequate source function control. At a linguistic level, speech can be viewed as organized by stress and intonation into suprasegmental units, with the segmental units nested within. The effect of the proposed teaching order might be to promote normal suprasegmental relations in speech, and thus to give the listener help in decoding the message. The second theme is that we teach "sounds" within a dynamic framework by rooting out, as a basis for training, the "posture and glide" view of speech. Ling says (p. 109), "The concept has probably taken a long time to die because tongue postures are typically shown in textbooks and because it is simple to think of phonemes as succeeding one another like letters on a printed page... The temptation to regard tongue postures as static rather than dynamic targets must be resisted, as must seduction by analogy with the written form." It is to the relationship between these themes that we wish to address ourselves. We know a little bit, though not very much, about the development of phonatory patterns in normal and deaf speakers. What is the relationship between suprasegmental and segmental patterns in normal development? Which faults are at the base of the unintelligibility of deaf speech?

THE DEVELOPMENT OF SUPRASEGMENTAL STRATEGIES IN NORMALS

Some years ago, Lieberman (1967) developed an ingenious theory of the origins of intonation in normal speech production. He pointed out that the intonation contour characteristic of most utterances may develop out of the relationship between respiration and phonation. In respiratory breathing, expiration and inspiration are of approximately equal duration, while in speech breathing, the expiratory limb occupies about 90% of the cycle. In phonation, the vocal folds must be brought into the air stream, as air is (typically) expelled fron the lungs. The duration over which air can be expelled provides a limit for the length of a phrase. In infants, as in adults, subglottal pressure typically drops at the end of a phonatory phrase (Bosma, Truby, & Lind, 1965). Since fundamental frequency of phonation is affected by subglottal pressure (van den Berg, 1958), one might expect a fall of fundamental frequency to mark the end of utterances. In a slightly different formulation, the "declination line" ('t Hart & Cohen, 1973)--a continuous fall in fundamental frequency throughout an utterance with decreasing f_0 of stress peaks as well as unstressed valleys--might be generated by falling subglottal pressure.

Objections have been raised to Lieberman's theory on several grounds. His assumption was that the mechanism for the pitch fall is a drop in subglottal pressure, but he substantially overestimated the numerical value of

the relationship between f_o and subglottal pressure; the values found by Baer (1979), 2 to 5 Hz/cmH_2O, have been obtained by a number of investigators (Fromkin & Ohala, 1968; Hixon, Klatt, & Mead, 1971). Since an inspection of data from Lieberman (1967), Collier (1975), and Atkinson (1973) suggests a pressure drop of 3 to 5 cmH_2O for sentences without emphatic stress, passive mechanisms would account for only about 15 Hz of the frequency drop. However, Maeda (1976) found f_0 drops of from 20 to 40 Hz in the productions of a corpus of sentences by an adult male. The discrepancy between f_0 fall and subglottal pressure drop suggests some kind of active use of the fundamental frequency contour to mark syntactic boundaries, especially since f_0 fall does not become greater, the longer the utterance (Breckenridge, Note 1). This result implies that speakers use a pitch look-ahead strategy in production, adjusting the slope of pitch fall to produce the same fall in long utterances as in short. Apparently, then, even if the pitch fall has its origin in respiratory dynamics, normal adult speakers use it in an active way to code syntactic information.

Durational manipulation is used by normal adult speakers in a strategy that is quite similar to that proposed for fundamental frequency—that is, to organize speech with respect to various syntactic considerations (Klatt, 1975). Vowels are longer in terminal positions in an utterance, and before clause boundaries. Vowel lengthening is also used to mark stressed syllables in an utterance. Indeed, a complex organization of interlocking duration rules is evidently used by a speaker to time syllables according to their position in words, clauses, or sentences, or with respect to the pattern of adjacent unstressed and stressed syllables (Lindblom & Rapp, 1973; Fowler, 1977). Tendencies of this sort of organization have been shown for many languages (Lehiste, 1970), although there are language-to-language differences in the way that apparently universal tendencies towards rhythmic performance organizations are applied (Allen, 1975).

Apparently, then, we may hypothesize that the primary aspects of acoustic production that mark the suprasegmentals--fundamental frequency of phonation and vowel duration--are organized by adult speakers over at least utterance lengths, using some kind of look-ahead strategy.

While this strategy has, perhaps, a physiological origin, it is clearly used actively by adult speakers to convey syntactic information about the structure of an utterance. If listeners depend on this information to decode segmental as well as suprasegmental information, then a failure to provide it might be an important source of the characteristic unintelligibility of deaf speech. In what follows, we review briefly what is known about suprasegmental production by deaf speakers.

THE DEVELOPMENT OF PITCH AND DURATION CONTROL IN PRELINGUALLY DEAF SPEAKERS

We know very little about the development of respiratory and phonatory control in deaf infants--indeed, we know very little about the development of such abilities in normal infants. Lenneberg, Rebelsky, and Nichols (1965) and Mavilya (1969) have suggested that in the first three to six months, the sounds produced by deaf and normal-hearing infants are much the same. Since phonatory output is affected by respiratory control, one might think that normal interaction between phonation and respiration was maintained in these

children; however, the phonatory output of the children was too incompletely studied for us to be at all confident of its characteristics. Furthermore, audiological assessment of the individual infants studied was inadequate to allow a specification of the degree of hearing loss. However, Stark (1972) has shown that spontaneous vocalizations of a group of severely-to profoundly-deaf children of ages sixteen to twenty-eight months are deviant in many ways. While some characteristics are like those of hearing children of a younger age, some are never seen in a normal population.

Older deaf children and deaf adults show deviant control of both respiration and phonation. As shown by Forner and Hixon (1977) and more recently by Whitehead (in press), the deaf do not use respiration in a way that would allow them to produce utterances of normal length, without great effort, due to a tendency to take in inadequate quantities of air, and to waste it during phonation. As to phonation, the most unintelligible deaf children show evidence of poor control of phonation, in the form of pitch breaks (McGarr & Osberger, 1978). Interestingly enough, in this last study, judgments of pitch deviancy were not very highly correlated with overall intelligibility.

The best documented aspect of deaf suprasegmental production is the overall deviance in timing. Deaf speakers produce fewer words per minute than normals, both because they prolong words and syllables, and because they pause for abnormal periods between words. They have, in addition, been reported to equalize the duration of stressed and unstressed syllables in sentence production, but careful studies by Reilly (1979) and Osberger (1978) show that this tendency is less salient than had been believed.

In spite of the deviancy of pitch and duration control in deaf speech, there is some evidence that the deaf may have some knowledge of the rules under which normals operate. In the paper we mentioned previously, Breckenridge (Note 1) asked two deaf speakers to produce sentences of different lengths. She found that, although their overall durations were grossly longer than those of normals, and they paused between words, there was, nonetheless, a tendency for a declination of pitch through the sentence. Reilly (1979) found that deaf speakers use duration to differentiate between stressed and unstressed syllables in disyllables (Figure 1), between primary and weak syllables in stress (Figure 2), and between prepausal and nonprepausal syllables (Figure 3). Thus, deaf speakers give evidence that they have some knowledge of the rules underlying suprasegmental production, even though their overall control of respiration and phonation is deviant. It is interesting to note that deaf speakers from traditional training programs probably were not explicitly taught these rules, since general discussions of suprasegmental organization in production are relatively recent in the research literature. Of course, whatever low frequency residual hearing the deaf possess could be used to recover fundamental frequency and durational information.

Before considering the effects of abnormally long durations on intelligibility, we should note an observation by Osberger (1978) that the prolongation often noted in deaf speech production is due chiefly to prolongation of vocalic segments. Such acoustic events as friction duration tend to be quite short. A somewhat similar observation has been made by Monsen (1976) for the voice onset time of voiceless stops. One might believe that an overall

prolongation of syllables would result in long voice onset times for voiceless consonants but, in fact, voice onset times are characteristically short. We believe that these "short" consonants result from incorrect programming of the timing of glottal and supraglottal events, that is, incorrect interarticulator programming, as we will discuss below.

Returning to the role of deviant suprasegmental production in generating unintelligible speech, there have been no studies, to our knowledge, of the role of pitch, but there have been attempts to assess the effect of training in the correction of timing, on the intelligibility of deaf speech. The results are inconsistent. John and Howarth (1965), and Heidinger (1972) found improvements in the intelligibility of the speech of children who were given training emphasizing timing, while Houde (Note 2) and Boothroyd, Nickerson, and Stevens (1974) did not. There is some indication that the changes in temporal control were accompanied by changes in other aspects of the children's speech. The other changes, whose patterns differed from one study to another, may have accounted for the intelligibility changes.

However, a better way to test the hypothesis that inappropriate timing is a significant contributor to the unintelligibility of deaf speech is through an analysis-by-synthesis approach; that is, by examining the perceptual effect of instrumental manipulation of recorded sentences. Following two somewhat preliminary studies using this approach (Lang, 1975; Bernstein, 1977), Osberger (1978) explored the effect in detail--and corrected pauses and vowel segment durations by adjusting the duration of the waveforms of simple sentences produced by deaf children. Her results are shown in Figure 4: Intelligibility scores averaged over six sentences, each produced by six children, as a function of six types of durational manipulation.

The manipulations were: correction (removal) of pauses, correction of relative timing, of stressed/unstressed syllable duration ratio, and of absolute duration. Corrections of pause and duration were also combined. The results show modest improvement for only one condition, the correction of relative timing. Whatever else may be said of the interpretation of this important experiment, it suggests that gross steady-state deviance of syllable duration in deaf speech is not a very large factor in the unintelligibility of the speech. This is not to say that durational information is not used by listeners in decoding speech. In Osberger's study, the sentences with the greatest number of perceived vowel errors showed the greatest improvement when relative duration was corrected, perhaps because segmental duration information about vowels became available. Indeed, studies of vowel duration production and perception in normals (Nooteboom, 1973) suggest that listeners are extremely sensitive to the duration that a vowel should have in a given context. Rather, Osberber's result suggests that for pronunciation that is grossly deviant, improvement of overall timing is insufficient to allow the listener to decode adequately. Neither does the result suggest that teaching children an appropriate suprasegmental strategy is unimportant. It has been shown by Calvert (1961) that experienced listeners to deaf speech cannot identify speech as deaf unless they hear at least syllable-length productions. Even if the sole effect of the characteristic deaf syllable prolongation were to make the deaf conspicuous and tedious to listen to, correction of deaf suprasegmental structure would still be a desirable training objective. But, as we will argue in the remainder of this paper, there are other aspects of

timing in deaf people's speech that are perceptually so deficient that they may be an important factor in the intelligibility of the speech, and therefore should be more fully investigated.

SOME EFFECTS OF ARTICULATORY TIMING ON SEGMENTAL DISTINCTIONS IN NORMALS

Some time ago, in addressing an earlier conference on the education of the deaf, our colleagues discussed the question of why speech spectrograms are hard to read (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1968). They described speech as a code, rather than a cipher. When listeners attempt to decode speech from an acoustic signal, they use cues that overlap along the time line--that is, they process information about one phone while they process information about another. This argument is related to that made by Ling, quoted earlier, that it is the dynamics of speech that have been neglected in deaf training. Recent perceptual experiments reinforce the importance of such dynamic factors in speech perception. We will not attempt a complete review but will stress vowel perception, because until recently, it was thought that vowels were identified largely through their static spectral characteristics. These static acoustic measures are associated with the central, quasi-steady-state region of syllables and correspond to the cavity resonances of static articulatory configurations. Strange, Verbrugge, Shankweiler, and Edman (1976) compared the identifiability of natural isolated vowels (which approximate sustained targets) and vowels spoken in a /pVp/ consonantal environment (where formant values are sustained minimally or not at all). Listeners identified the vowels with greater accuracy when listening to the /pVp/ syllables, indicating that sustained target information is not a sufficient condition for the highest accuracy to be achieved. Figure 5, from an unpublished paper of this same group, summarizes the results of several related experiments. Vowels were produced by two speakers, and synthesized by an OVE synthesizer, in three types of context. Condition "Natural" had sustained natural vowels. In condition "Steady State" a few pitch periods from a sustained vowel were iterated, using computer techniques, to produce a syllable equal in duration to the matching natural vowel, with no formant movement at all. A CVC context /bVb/ was also used. Clearly, CVC syllables are better identified than isolated vowels, but natural isolated vowels are more identifiable than iterated vowels. Apparently, even the minimal dynamics of a natural isolated vowel carry segmental information, but sustained "target" values are relatively uninformative.

When we turn to the consonants, there is much information indicating the importance of dynamic parameters in segment identification. For example, Figure 6 shows the results of an early experiment on the discrimination of stops, semivowels and diphthongs (Liberman, Delattre, Gerstman, & Cooper, 1956). Their results show that a synthetic CV syllable with a transition of short duration will be perceived as beginning with a stop. If the transition is made somewhat longer, the syllable will be perceived as beginning with a semivowel, while if it is made longer still, the syllable is perceived as a vowel of changing color. Indeed, it is clear, as Miller (in press) has observed, that "prosodic and segmental information is so intertwined in speech that only by studying both factors will it ultimately be possible to build a model of segmental perception."

In production, dynamic information is generated by the movement of the articulators. Changes in dynamics can be generated by changing the speed of movement of individual articulators, or by changing the relative timing. It is this latter type of change in articulator dynamics that is of great interest to us.

The best known example of interarticulator programming is voice onset time. Figure 7 shows the distribution of voice onset times for /p/ and /b/ in English. It is well known that the timing of the release of oral occlusion relative to the onset of glottal pulsing distinguishes the voiced from the voiceless stop (Lisker & Abramson, 1967). It can also be shown that the difference between aspirated stops seems to be one of articulatory timing (Löfqvist, in press). Aspects of laryngeal-supralaryngeal timing have been most widely investigated, but one can find examples in the literature that suggest that the relative time at which articulators move is crucial in other ways in producing segmental information. For example, Kent and Moll (1975) used cinefluorography to investigate the articulation of consonant clusters beginning with /sp--/ and found that the closure for /p/ and release of the constriction for /s/ occurred almost simultaneously, irrespective of linguistic environment. Recent work of our own suggests that when syllables are produced with varying stress or speaking rate, although there are changes in vowel duration and target vocal tract shape, the relative time of onset of movement (as inferred from EMG measures) for tongue fronting and lip closure varies relatively little. It is our argument that a failure to control interarticulator programming contributes substantially to the unintelligibility of deaf speech.

ARTICULATORY TIMING IN THE SPEECH OF THE DEAF

Interarticulator programming has not been much examined in considering the aberrant qualities of deaf speech, yet it is possible to find many hints in the literature of the importance of this variable. We mentioned above that Monsen (1976) has shown a lack of voice-onset-time distinctions in deaf speech production. He also noted (1978) that intelligibility was well correlated with the extent of F_2 transition in the production of the syllable /ais/. Of course, an acoustic transition in a spectrogram is the result of several articulatory events. In the case given, the transition might appear to be abnormal either if vowels were not correctly placed (and it is well known that the deaf usually show a reduced vowel space) or if the tongue and jaw movements for the diphthong were not correctly coordinated.

It is our belief that temporal coordination, rather than absolute articulator placement, deserves more investigation than it has thus far received. Let us indicate how such investigations might proceed. In the example we will show, we stress an EMG approach. Figure 8 shows a schematic EMG representation of the syllable /pip/ (Bell-Berti, in press). For this syllable, orbicularis oris (OO) activity is associated with the initial and terminal /p/. This muscles purses and closes the lips. For the vowel /i/, the genioglossus (GG) muscle bunches the tongue and brings it forward in the mouth. These events have a normal time relationship, as shown in (a). When the syllable is prolonged, as when stress is changed, either of two things can happen. One possibility is that the duration of the articulatory events can change, as shown in (b). Another possibility is that the overlap between

articulatory events can change, as shown in (c). (Of course, both could happen together.) We believe that in normals, mechanism (b) is heavily used. In the deaf, is interarticulator timing is less well controlled, inappropriate overlap or changes may occur. We believe that errors of this sort contribute heavily to the unintelligibility of deaf speech.

We have done a pilot experiment that demonstrates such a failure of interarticulator programming. In this study, acoustic and electromyographic measures were made of an adult deaf speaker and a hearing speaker. The prelingually deafened speaker (mean pure tone average 150 dB+ISO) is a graduate of an oral school of the deaf and has received remedial speech classes as an adult. The hearing speaker has frequently served as a subject for EMG experiments.

Each subject produced many repetitions of each of several utterance types, including /əpipɑp/ with stress on /i/ and or /ɑ/. Surface electrodes were used in recording from the orbicularis oris, and conventional hooked wire electrodes were inserted into the genioglossus. Electromyographic data were analyzed by previously described techniques (Kewley-Port, 1973). All the utterances analyzed were intelligible to listeners.

Some rather typical data for the hearing speaker are shown in Figure 9, the utterance type /ə́pipɑp/. At the top of each column--either GG or OO--is the ensemble average obtained by averaging all the repetitions of the utterance type; four single tokens are seen in the columns below. For each utterance type, the line-up point for EMG and acoustic events is indicated by the dashed line at Ø msec; the acoustic event at that point is the closure release of the first /p/.

Looking first at the GG column, we find that peak activity is higher for /i/ than /ɑ/, as would be expected. Indeed, peak GG activity for the vowel occurs approximately at the time of the acoustic line-up--the /p/ burst release event. Turning to the OO column we find three well-defined peaks corresponding to the lip movement for the /p/ gestures, with the line-up falling between peaks 1 and 2. A striking feature of these data is the similarity of the EMG patterns observed for the individual tokens, for both muscles.

Figure 10 shows data for the utterance /əpí́pɑp/, by the hearing speaker. The duration of the genioglossus activity is shorter in this utterance type, and the time between the peaks of OO activity is shifted. Both of these changes are direct effects of the shortening of /i/ as stress is shifted to /ɑ/. Another effect of de-stressing /i/ is the decrease in average peak height. All these characteristics of the effects of stress change have been previously noted (Harris, 1978). Note, however, that the pattern of peak activity for GG still occurs simultaneously with the /p/ release.

Figure 11 shows the data for the deaf subject's production of /ə́pipɑp/. First, examining averaged EMG activity for OO, we see that as in data for the hearing subject, there are three well-defined peaks of activity. Duration is prolonged overall for the deaf speaker, and /p/ release falls essentially between the peaks 1 and 2 as we observed for the hearing speaker. Turning to GG, peak activity is less well defined and appears later relative to burst

release, compared to the hearing speaker's pattern. That is, peak activity occurs after the deaf speaker has produced the acoustic /p/ release.

In other words, activity for lip closure for the consonant and tongue bunching for the vowel overlap more in the hearing subject than in the deaf speaker. Further, there is considerable variability in duration of GG activity from token to token. In some instances, this activity starts fairly early (e.g., Token 2) and at other times, later (e.g., Token 1). Token-to-token variability is far more striking for GG than for OO.

Figure 12 shows the data for the deaf speaker's production of /əpípap/. The same comments may be made as for the previous figure.

The results of the study suggest that the deaf speaker may have appropriate control of a visible articulator, the lips, but does not control the movement of the tongue so consistently, with a resulting inability to program the two articulators with respect to each other. Such inconsistencies would have substantial effects on the acoustic signal. For example, the duration of the transition depends on the coordination of lips, tongue and jaw, and the generation of the stop burst depends on the rapid release of an occlusion by the coordinated activity of the articulators, after an appropriate intra-oral pressure has been built up.

A failure of interarticulator coordination, then, affects the dynamic character of the speech signal in ways that are difficult to characterize as strictly "segmental" or "suprasegmental." We believe that a consideration of the acoustic consequences of such failures may be an important contributor to the unintelligibility of deaf speech.

REFERENCE NOTES

1. Breckenridge, J. Declination as a phonological process. Unpublished manuscript, 1977. (Bell Laboratories.)
2. Houde, R. Instantaneous visual feedback in speech training for the deaf. Paper presented at the American Speech and Hearing Association Convention, Detroit, 1973.

REFERENCES

Allen, G. D. Speech rhythm: Its relation to performance universals and articulatory timing. Journal of Phonetics, 1975, 3, 75-86.

Atkinson, J. E. Aspects of intonation in speech: Implications from an experimental study of fundamental frequency. Unpublished doctoral dissertation, University of Connecticut, 1973.

Baer, T. Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. Journal of the Acoustical Society of America, 1979, 65, 1271-1275.

Bell-Berti, F. Inter-articulator programming and clinical advice to "slow down." Journal of Childhood Communication Disorders, in press.

Bernstein, J. Intelligibility and simulated deaf-like speech. Conference Record (IEEE International Conference on Acoustics, Speech and Signal Processing, Hartford, Conn.), 1977.

Boothroyd, A., Nickerson, R., & Stevens, K. Temporal patterns in the speech

of the deaf--A study in remedial training. Northampton, Mass.: C. V. Hudgins Diagnostic and Research Center, Clark School for the Deaf, 1974.

Bosma, J. F., Truby, H. M., & Lind, J. Cry motions of the newborn infant. In J. Lind (Ed.), Newborn infant cry. Uppsala: Almqvist and Wiksell, 1965.

Calvert, D. Some acoustic characteristics of the speech of profoundly deaf individuals. Unpublished doctoral dissertation, Stanford University, 1961.

Collier, R. Physiological correlates of intonation patterns. Journal of the Acoustical Society of America, 1975, 58, 249-255.

Forner, L. L., & Hixon, T. J. Respiratory kinematics in profoundly hearing-impaired speakers. Journal of Speech and Hearing Research, 1977, 20, 373-497.

Fowler, C. Timing control in speech production. Unpublished doctoral dissertation, University of Connecticut, 1977.

Fromkin, V. A., & Ohala, J. Laryngeal control and a model of speech production. Working Papers in Phonetics, University of California at Los Angeles, 1968, 10, 98-110.

Harris, K. S. Vowel duration change and its underlying physiological mechanisms. Language and Speech, 1978, 21, 354-361.

Heidinger, V. A. An exploratory study of procedures for improving temporal patterns in the speech of deaf children. Unpublished doctoral dissertation, Teachers College, Columbia University, 1972.

t'Hart, J., & Cohen, A. Intonation by rule; a perceptual quest. Journal of Phonetics, 1973, 1, 309-327.

Hixon, T. J., Klatt, D., & Mead, J. Influence of forced transglottal pressure change on vocal fundamental frequency. Journal of the Acoustical Society of America, 1971, 49, 105(A).

John, J. D. J., & Howarth, N. J. The effect of time distortions on the intelligibility of deaf children's speech. Language and Speech, 1965, 8, 127-134.

Kent, R. D., & Moll, K. L. Articulatory timing in selected consonant sequences. Brain and Language, 1975, 2, 304-323.

Kewley-Port, D. Computer processing of EMG signals at Haskins Laboratories. Haskins Laboratories Status Report on Speech Research, 1973, SR-33, 173-183.

Klatt, D. Vowel lengthening is syntactically determined in a connected discourse. Journal of Phonetics, 1975, 3, 129-140.

Lang, H. G. A computer based analysis of the effects of rhythm modification on the intelligibility of the speech of hearing and deaf subjects. Unpublished master's thesis, Rochester Institute of Technology, 1975.

Lehiste, I. Suprasegmentals. Cambridge, Mass.: M.I.T. Press, 1970.

Lenneberg, E. H., Rebelsky, F. G., & Nichols, I. A. The vocalizations of infants born to deaf and hearing parents. Human Development, 1965, 8, 23-37.

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. Why are speech spectrograms hard to read? American Annals of the Deaf, 1968, 113, 127-133.

Liberman, A. M., Delattre, P., Gerstman, L., & Cooper, F. S. Tempo of frequency change as a cue for distinguishing classes of speech sounds. Journal of Experimental Psychology, 1956, 52, 127-137.

Lieberman, P. Intonation, perception and language. Cambridge, Mass.: M.I.T. Press, 1967.

Lindblom, B. E. F., & Rapp, K. Some temporal regularities of spoken Swedish. Papers from the Institute of Linguistics (University of Stockholm), 1973, 21, 1-59.

Ling, D. Speech and the hearing-impaired child: Theory and practice. Washington, D.C.: A. G. Bell, 1976.

Lisker, L., & Abramson, A. S. Some effects of context on voice onset time in English stops. Language and Speech, 1967, 10, 1-28.

Löfqvist, A. Interarticulator programming in stop production. Journal of Phonetics, in press.

Maeda, S. A characterization of American English intonation. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1976.

Mavilya, M. P. Spontaneous vocalization and babbling in hearing-impaired infants. Unpublished doctoral dissertation, Teachers College, Columbia University, 1969.

McGarr, N. S., & Osberger, M. J. Pitch deviancy and intelligibility of deaf speech. Journal of Communication Disorders, 1978, 11, 237-247.

Miller, J. The effect of speaking rate on segmental distinctions. In P. D. Eimas and J. Miller (Eds.), Perspectives in the study of speech. Hillsdale, N.J.: Erlbaum, in press.

Monsen, R. B. The production of English stop consonants in the speech of deaf children. Journal of Phonetics, 1976, 4, 29-41.

Monsen, R. B. Toward measuring how well deaf children speak. Journal of Speech and Hearing Research, 1978, 21, 197-219.

Nooteboom, S. C. The perceptual reality of some prosodic durations. Journal of Phonetics, 1973, 1, 24-45.

Osberger, M. J. The effect of timing errors on the intelligibility of deaf children's speech. Unpublished doctoral dissertation, City University of New York, 1978.

Reilly, A. P. Syllable nucleus duration in the speech of hearing and deaf children. Unpublished doctoral dissertation, City University of New York, 1979.

Stark, R. Some features of the vocalizations of young deaf children. In J. F. Bosma (Ed.), Third symposium on oral sensation and perception. Springfield, Ill.: Charles C. Thomas, 1972, 431-446.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. Consonant environment specifies vowel identity. Journal of the Acoustical Society of America, 1976, 60, 213-224.

van den Berg, J. Myoelastic aerodynamic theory of voice production. Journal of Speech and Hearing Research, 1958, 1, 227-244.

Whitehead, R. Some respiratory and aerodynamic patterns in the speech of the hearing impaired. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired: Research, training and personnel preparation. Washington, D.C.: A. G. Bell Association, in press.

Figure Legends

Figure 1. Syllable nucleus durations in disyllables. Mean values for 16 congenitally deaf students, 16 normal-hearing children, and a highly intelligible deaf subject. (After Reilly, op. cit.)

Figure 2. Mean syllable nucleus durations for primary and weak stress syllables; groups as described above (Reilly, op. cit.).

Figure 3. Mean syllable nucleus durations in prepausal and non-prepausal syllables; groups as described above (Reilly, op. cit.).

Figure 4. Intelligibility scores of deaf utterances averaged across six subjects and six sentences for six types of durational manipulation: (1) sentences unaltered; (2) pauses corrected; (3) relative timing corrected; (4) absolute syllable duration corrected; (5) relative syllable duration and pauses corrected; (6) absolute syllable duration and pauses corrected. (After Osberger, op. cit.)

Figure 5. Percent identification errors for vowels as sampled from the speech of two speakers, and synthesized. "Natural" vowels were produced as isolated vowels. "Steady-state" vowels were produced by iterating two pitch periods. "CVC" syllables were produced in the context /bVb/.

Figure 6. Distribution of identification by subjects presented with synthetic stimuli with the transition durations indicated below. (After Liberman et al., op. cit.)

Figure 7. Distribution of voice onset time for stops in isolated words. (After Lisker and Abramson, op. cit.)

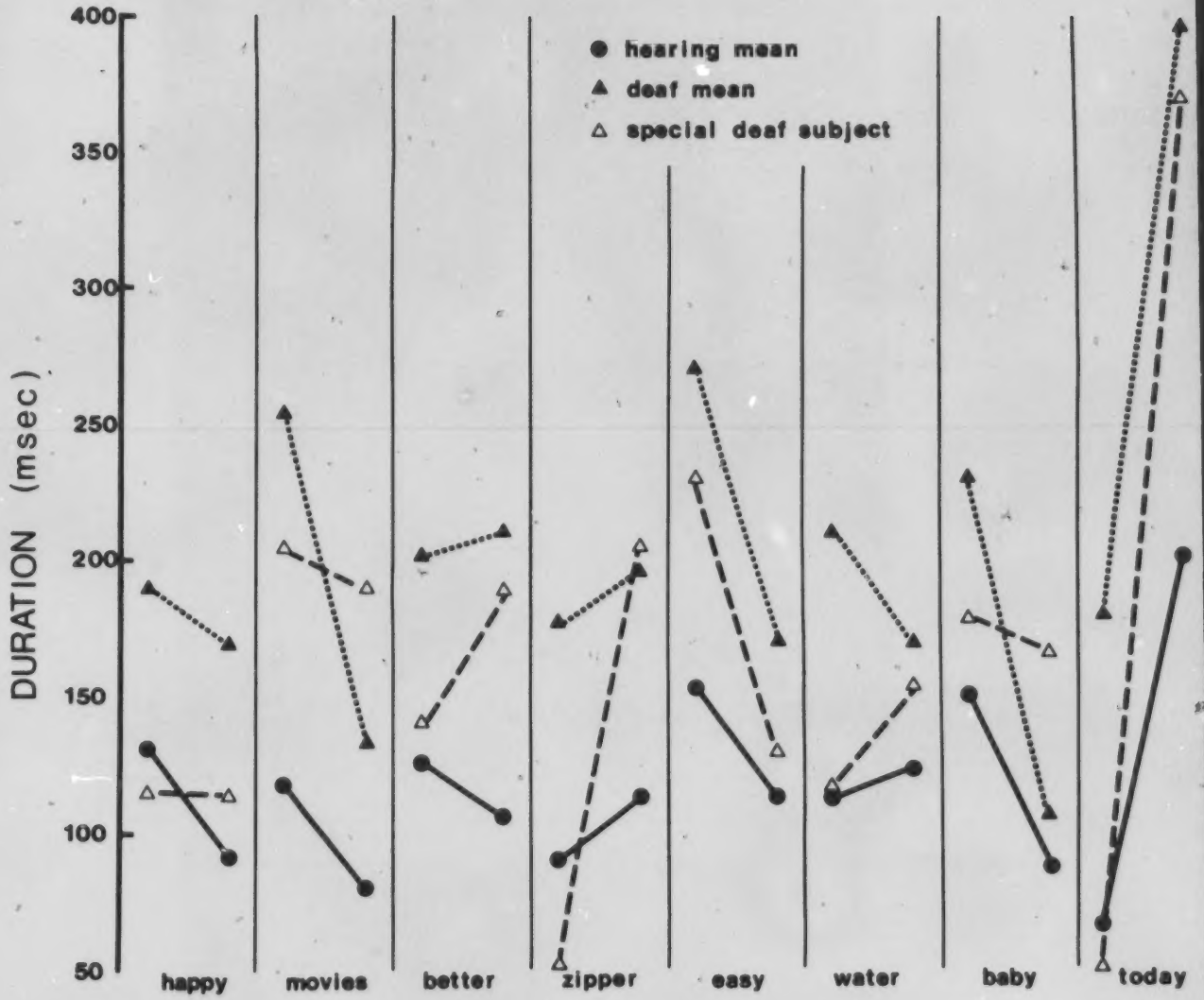
Figure 8. Hypothetical dynamics of /pip/ production of (a) normal duration and (b) and (c) lengthened duration. (After Bell-Berti, in press.)

Figure 9. [ə'pi pɑp] as produced by a normal speaker. Data plots at the top show EMG averaged for about 16 tokens for the genioglossus and orbicularis oris muscles. Four individual tokens for each muscle are shown below. Dotted line indicates acoustic release of [p] closure.

Figure 10. [əpi'pɑp] as produced by a normal speaker. Data presented as in Figure 9.

Figure 11. [ə'pi pɑp] as produced by a deaf speaker. Data presented as in Figure 9.

Figure 12. [əpi'pɑp] as produced by a deaf speaker. Data presented as in Figure 9.


400
350
300
250
200
150
100
50
DURATION (msec)
hearing mean
deaf mean
special deaf subject
happy
movies
better
zipper
easy
water
baby
today

igure 1

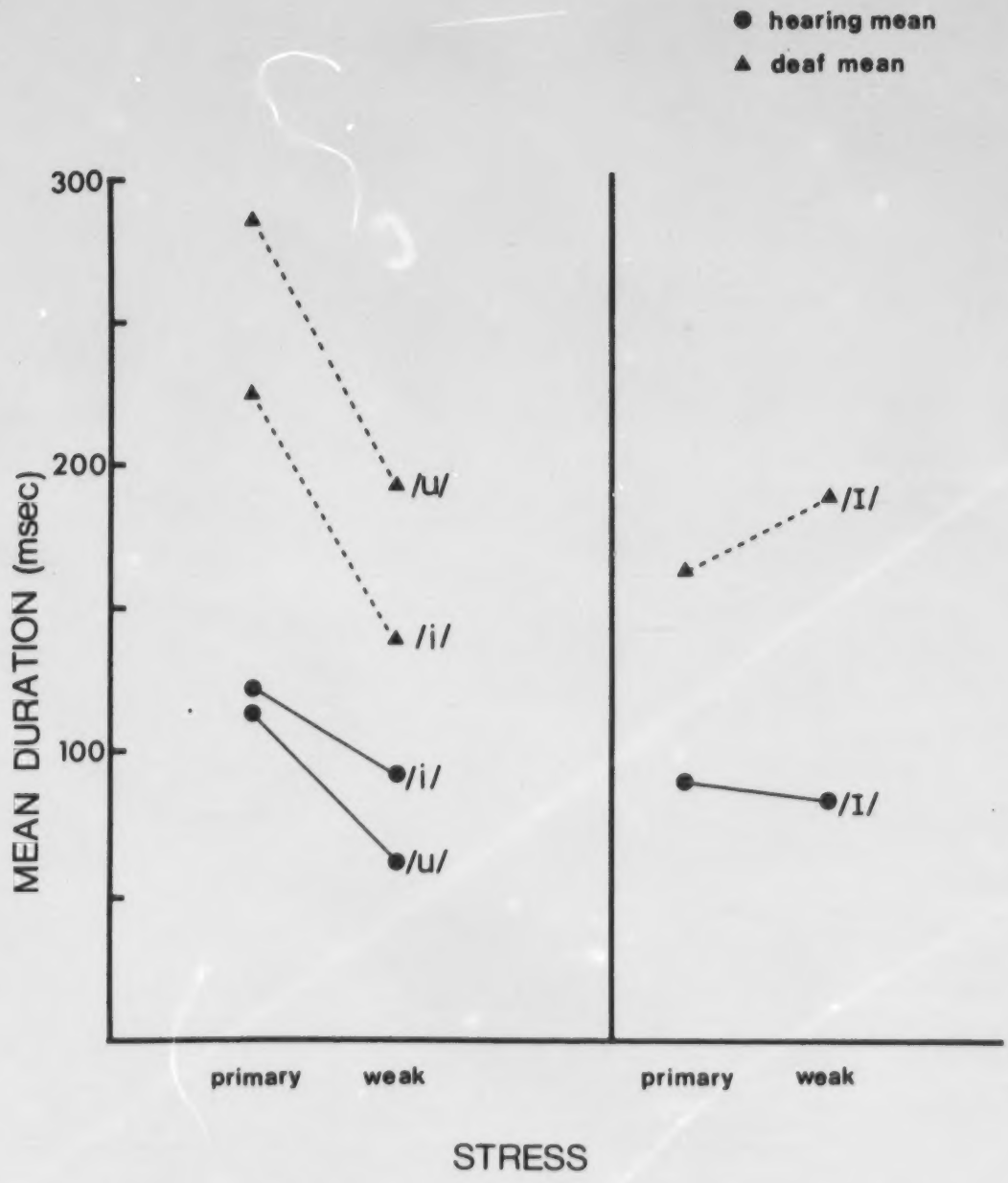

hearing mean
deaf mean
300
200
100
MEAN DURATION (msec)
/u/
/i/
/i/
/u/
/I/
/I/
primary
weak
primary
weak
STRESS

Figure 2

11

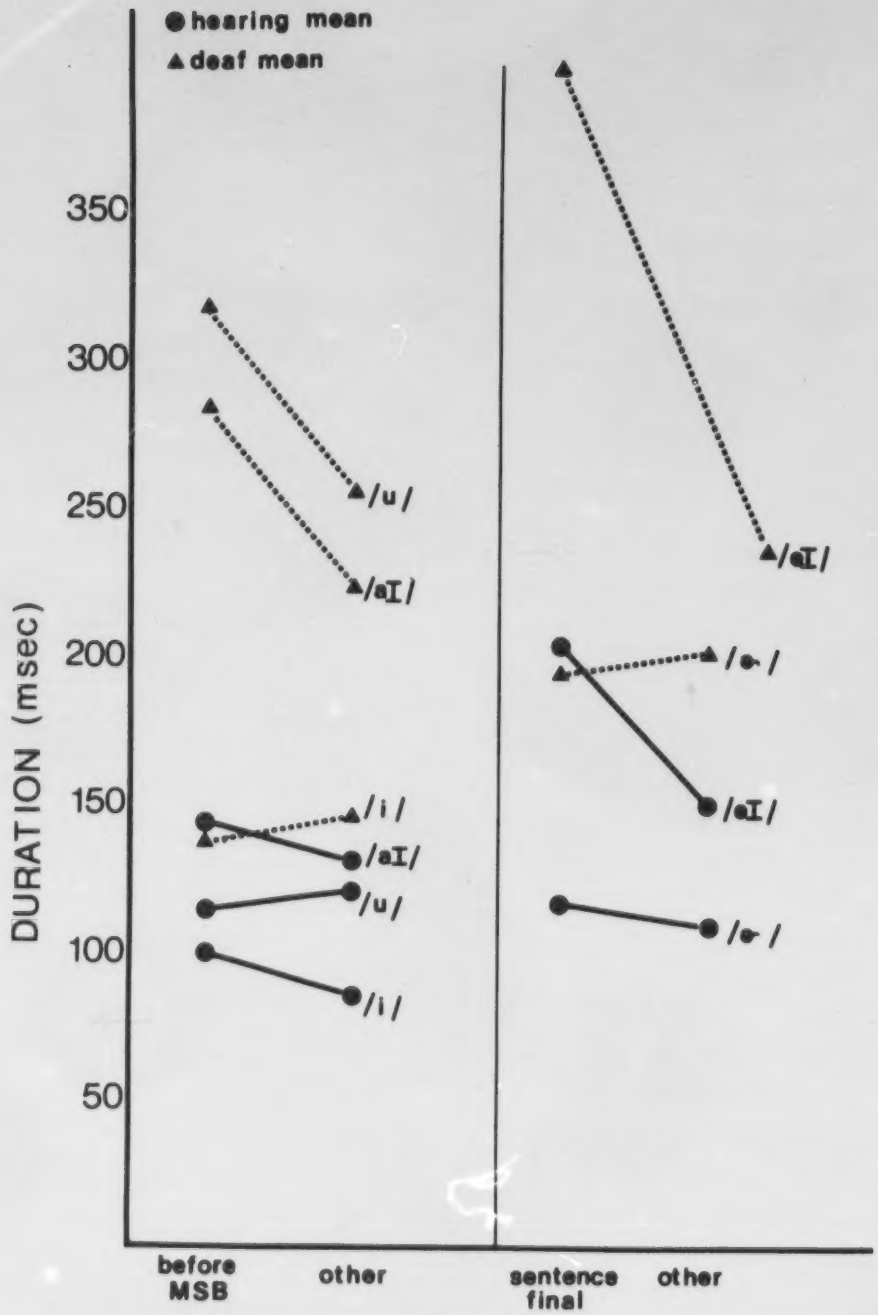
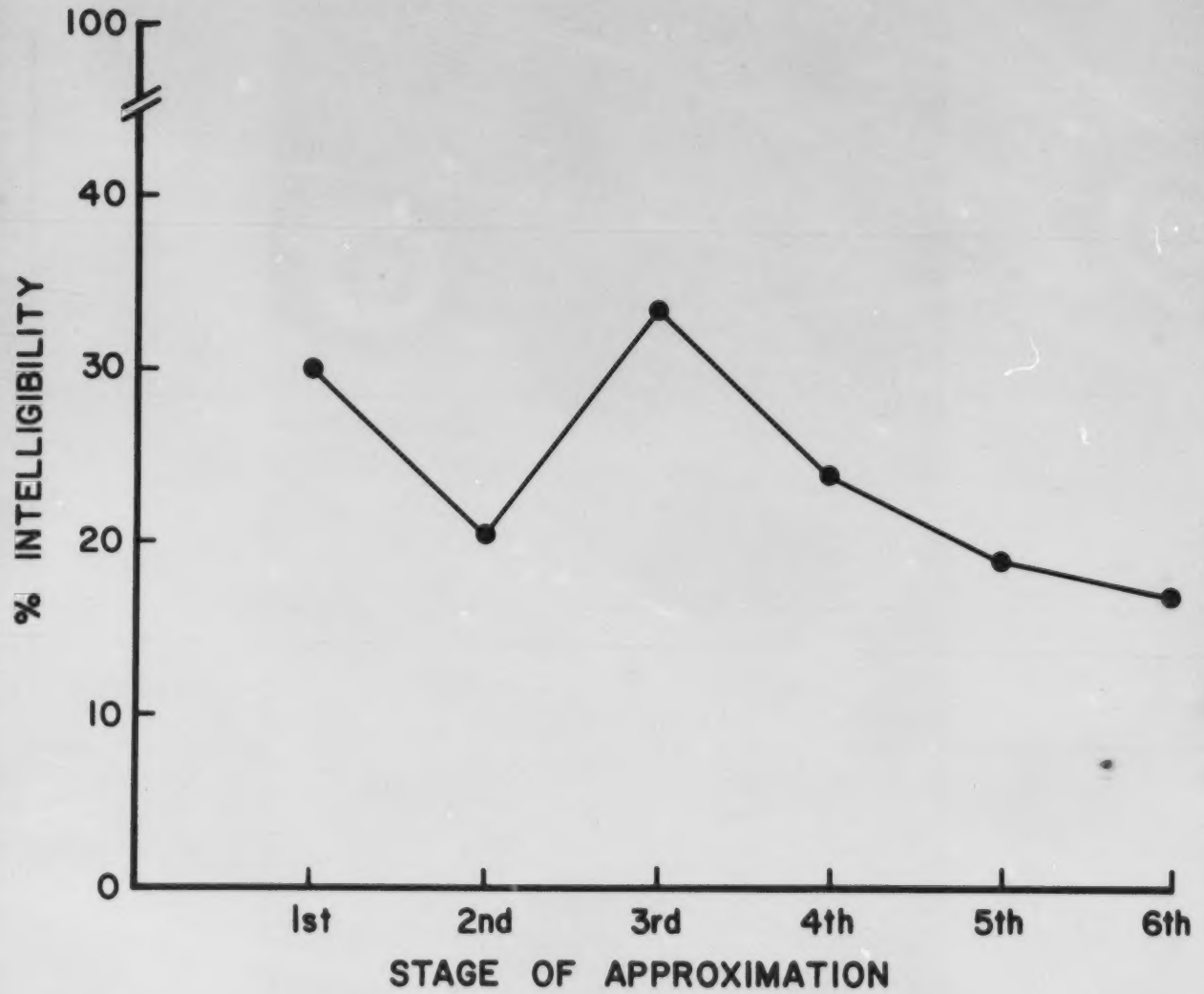

hearing mean
deaf mean
DURATION (msec)
350
300
250
200
150
100
50
/u/
/aI/
/i/
/aI/
/u/
/i/
/aI/
/ɔ/
/aI/
/ɔ/
before MSB
other
sentence final
other

Figure 3

Figure 4


100
40
30
20
10
0
% INTELLIGIBILITY
1st
2nd
3rd
4th
5th
6th
STAGE OF APPROXIMATION

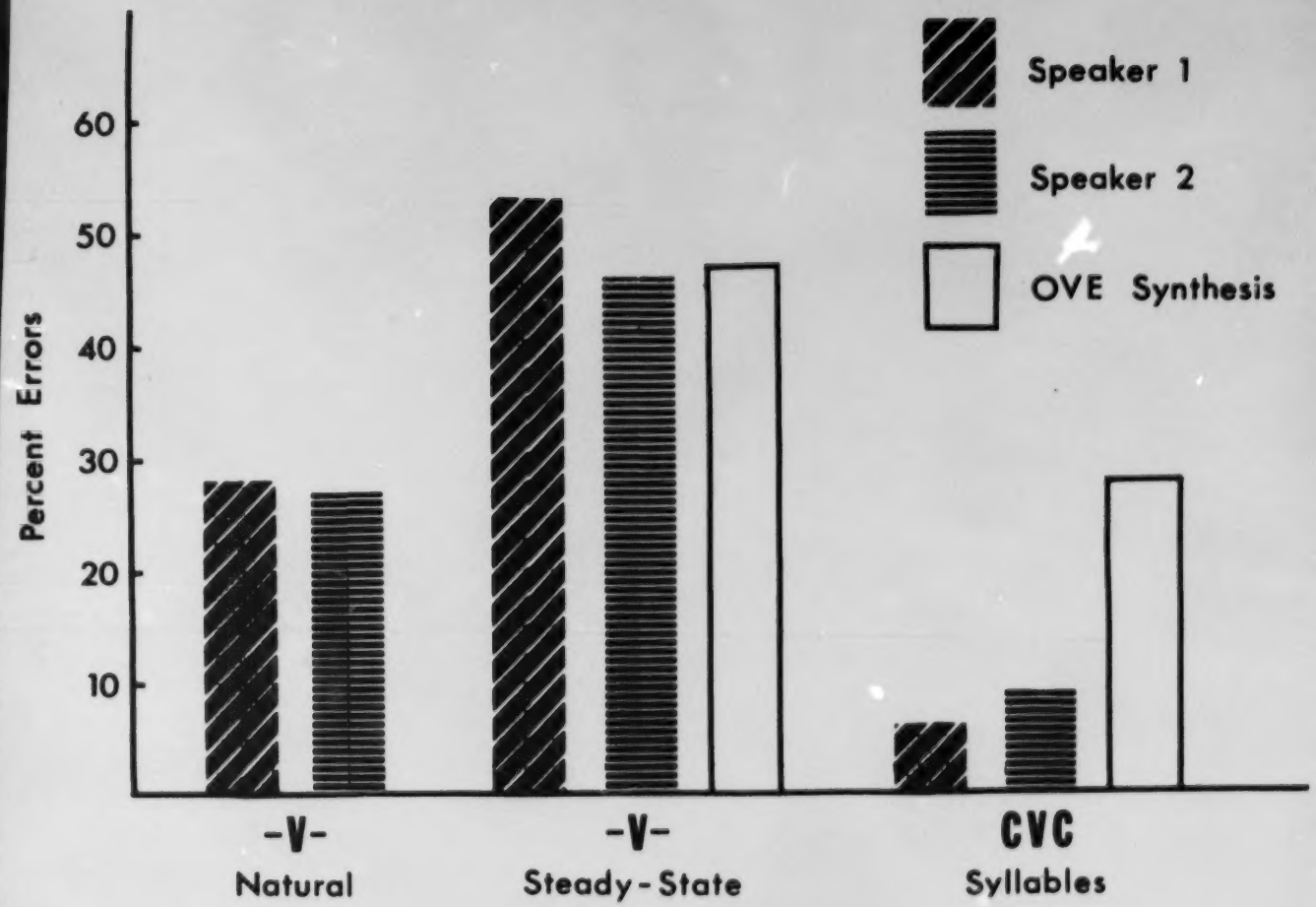

Percent Errors
60
50
40
30
20
10
Speaker 1
Speaker 2
OVE Synthesis
-V-
Natural
-V-
Steady-State
CVC
Syllables

Figure 5

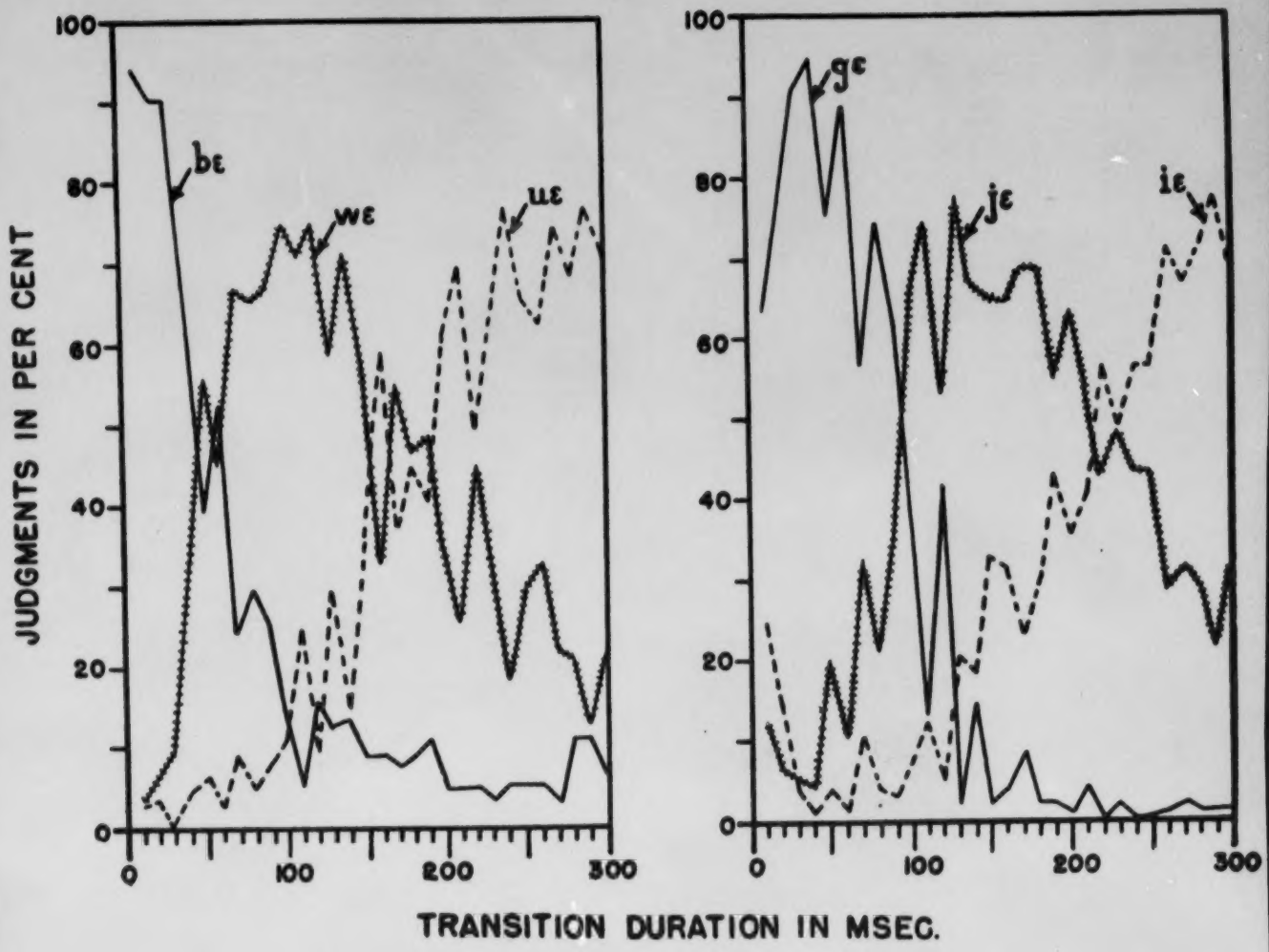

100
80
60
40
20
0
JUDGMENTS IN PER CENT
bɛ
wɛ
uɛ
0
100
200
300
gɛ
jɛ
iɛ
TRANSITION DURATION IN MSEC.

Figure 6

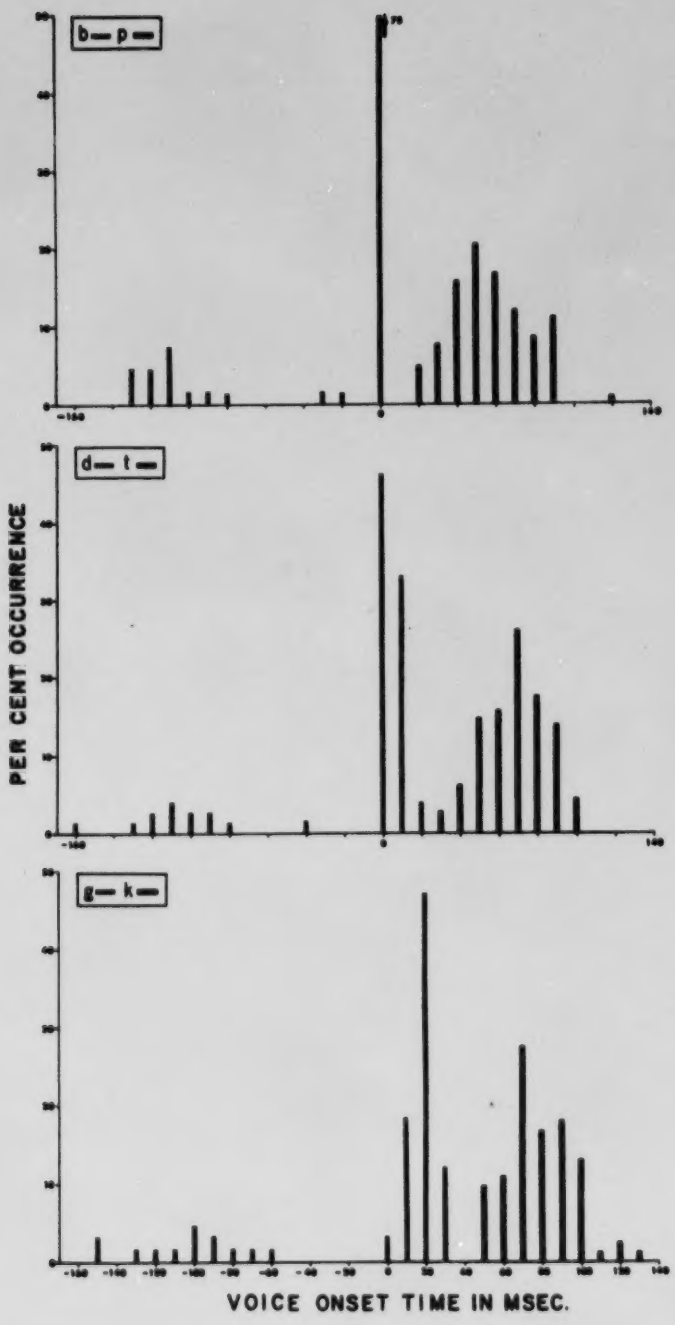
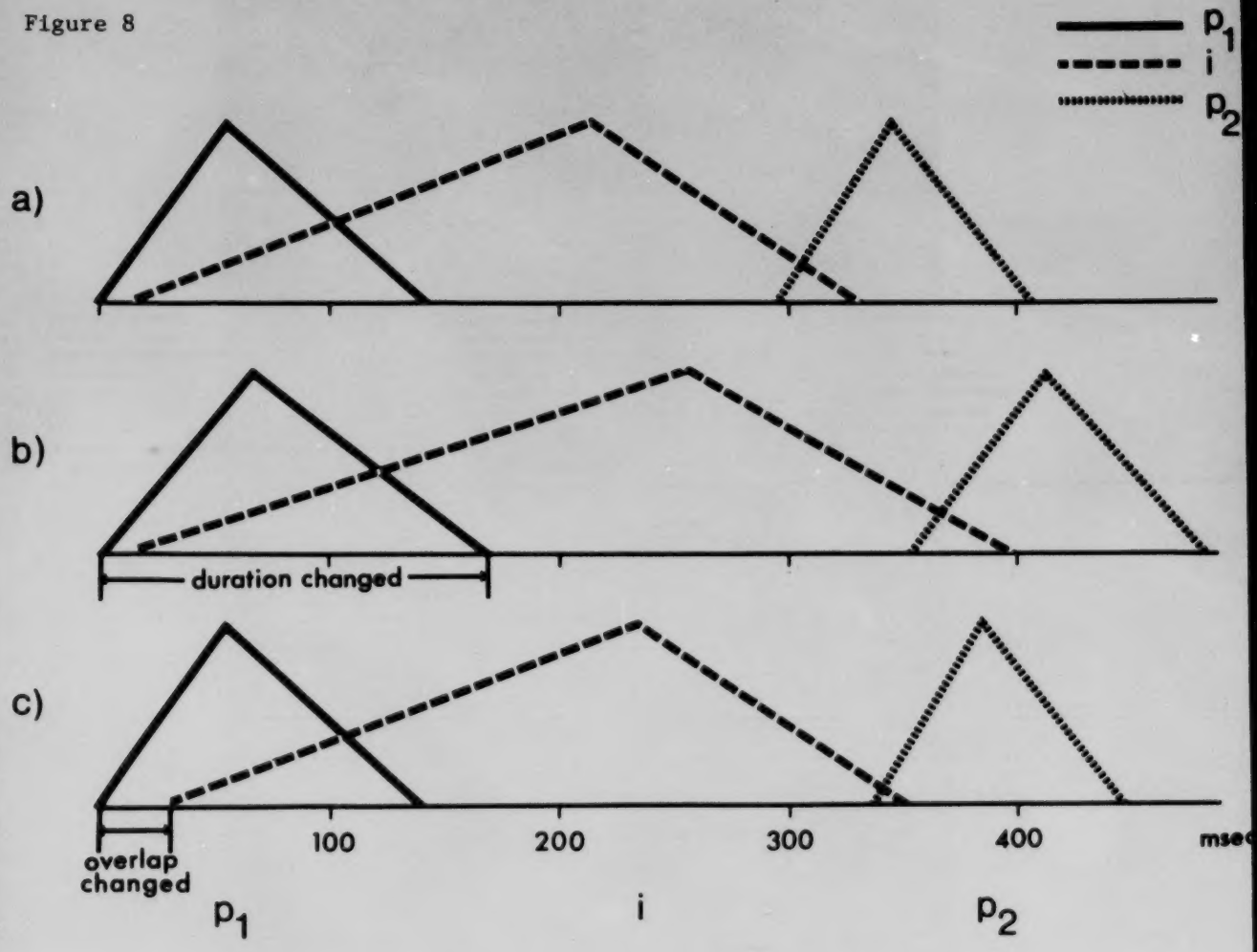


Figure 7

Figure 8


p1
i
p2
a)
b)
duration changed
c)
overlap changed
100
200
300
400
msec
p1
i
p2

[ə'pipɑp]

FBB

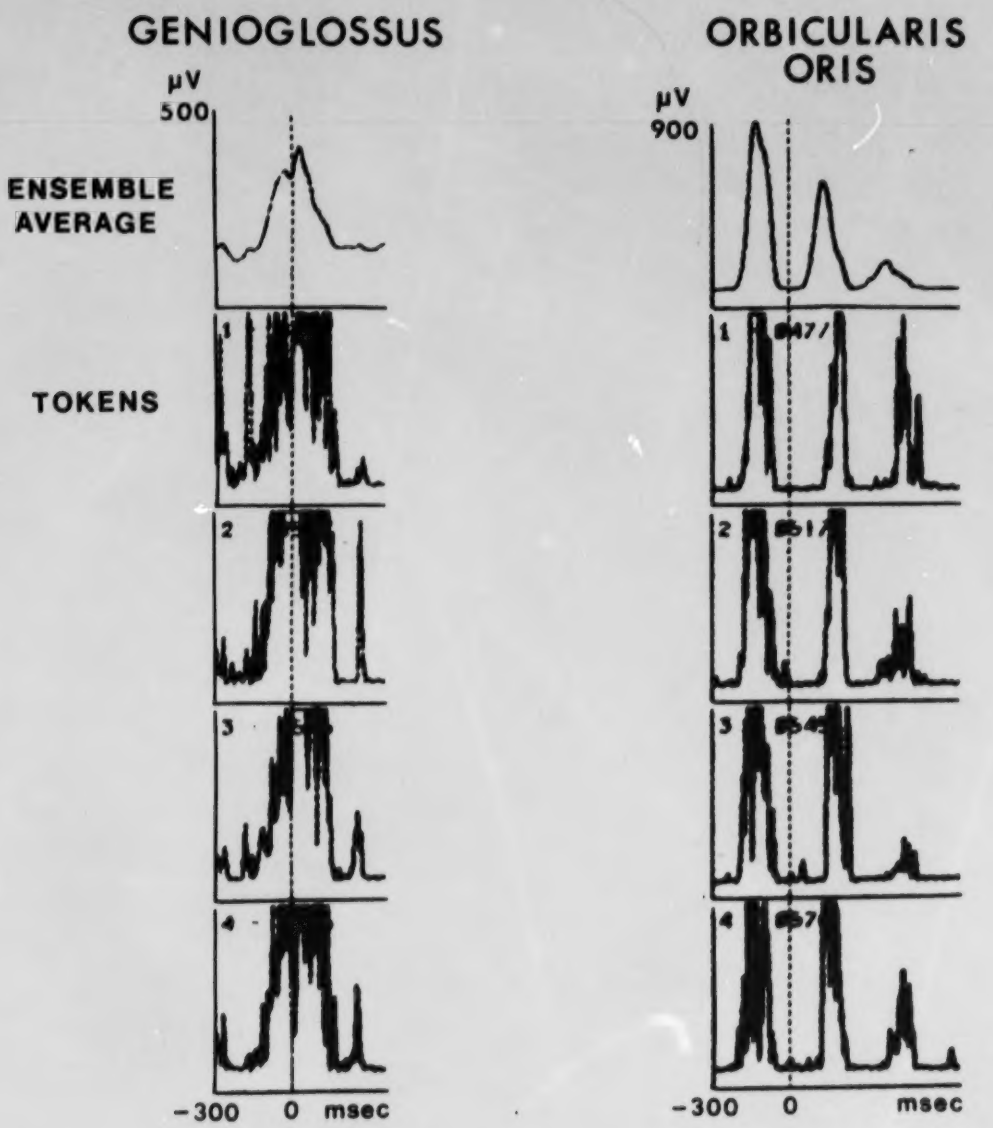


Figure 9

[əpi'pap]

FBB

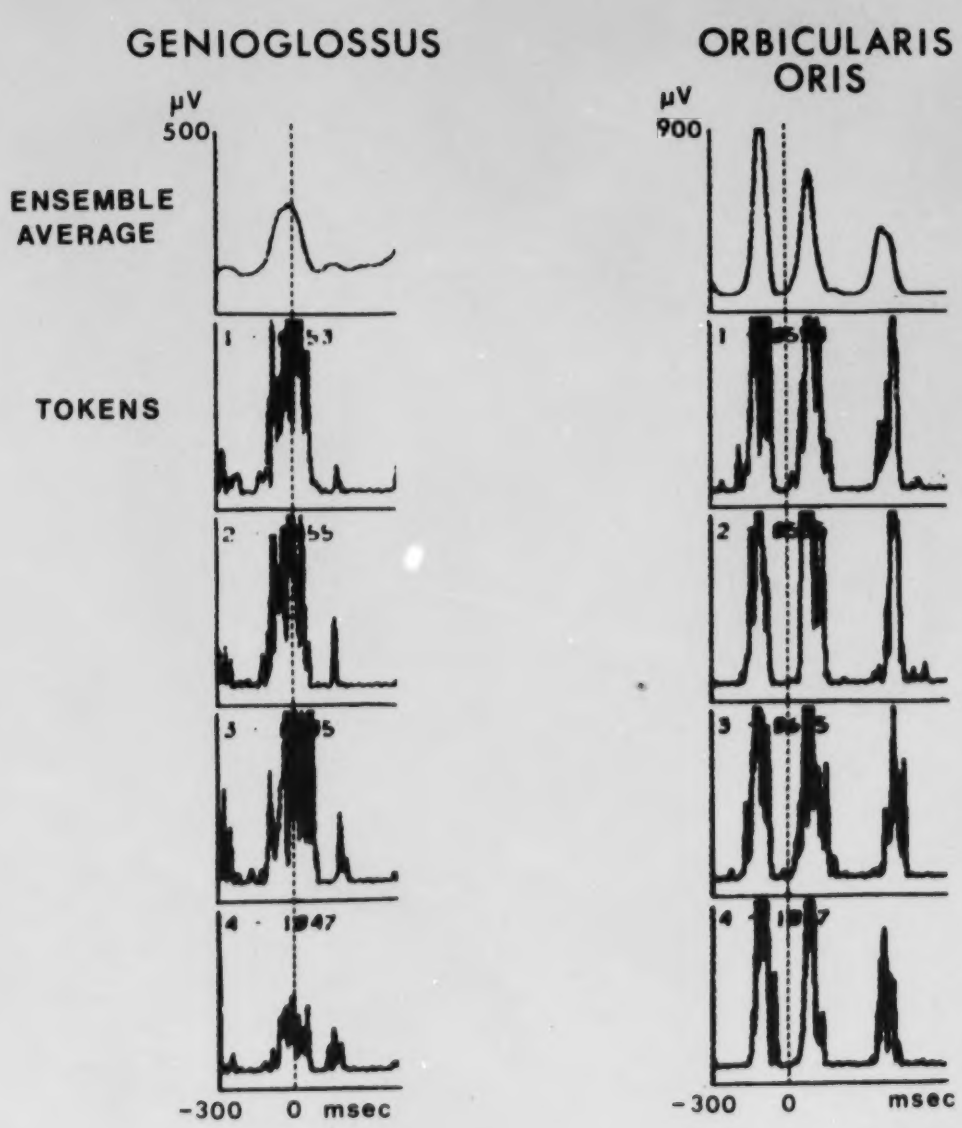


Figure 10

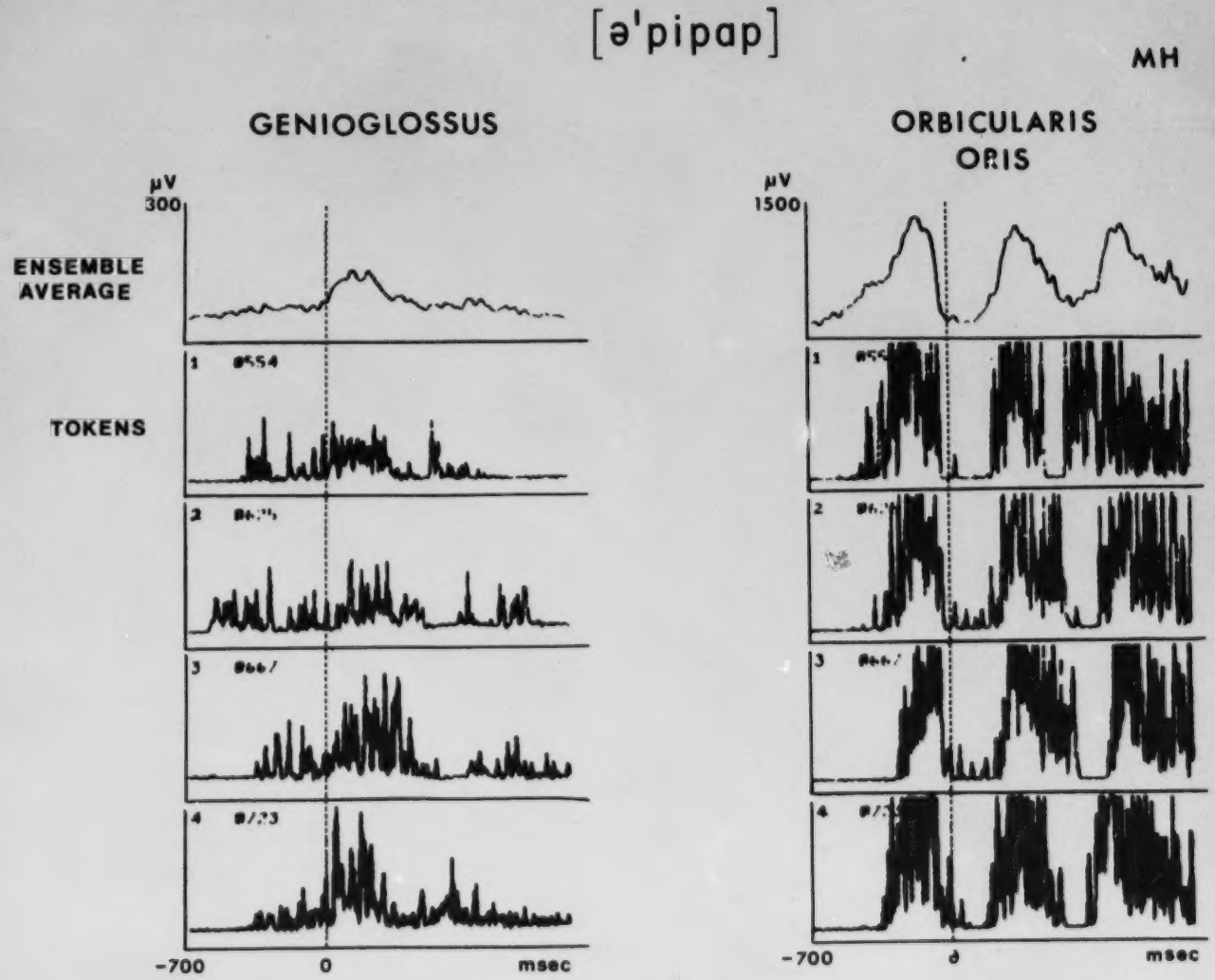
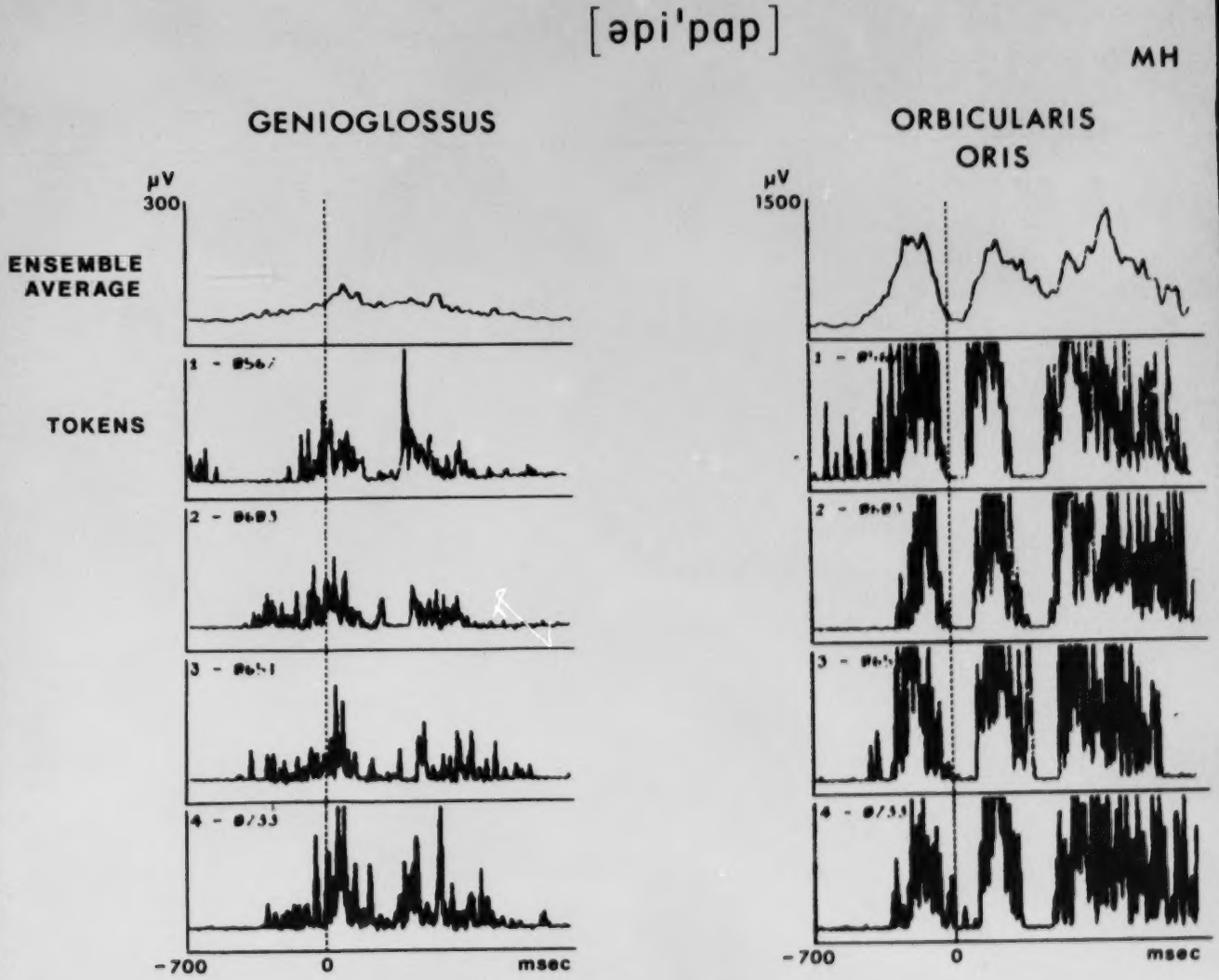

[ə'pipap]
MH
GENIOGLOSSUS
ORBICULARIS ORIS
μV
300
ENSEMBLE AVERAGE
TOKENS
1
2
3
4
-700
0
msec
μV
1500
-700
msec

Figure 11

Figure 12


[əpi'pɑp]
MH
GENIOGLOSSUS
ORBICULARIS ORIS
μV
300
1500
ENSEMBLE AVERAGE
TOKENS
-700
0
msec

ACCESSIBILITY OF THE VOICING DISTINCTION FOR LEARNING PHONOLOGICAL RULES*

Alice F. Healy+ and Andrea G. Levitt++

Abstract. In three experiments, a series of nonsense syllables ending in consonants was presented to adult subjects who had to discover or learn a rule for classifying them into two groups. The rule was based either on the voicing of the final consonants or on an arbitrary division of them. Subjects performed better with the voicing than with the arbitrary rule only when there was a straightforward relationship between the voicing rule and the plural formation rule in English or, more generally, when voicing assimilation with an added consonant was involved and attention was focused on the sound and articulation of the syllables. We conclude that the voicing distinction is not ordinarily accessible and that individuals easily learn and use phonological rules involving voicing assimilation because of articulatory constraints on the production of consonant clusters.

In order to describe the sounds of the English language, linguists have adopted various systems of distinctive features (see, for example, Jakobson, Fant, & Halle, 1963). One such distinctive feature corresponds to the presence or absence of voicing. Voiced consonants -- such as /b, d, g, v, z/ -- involve vibrations of the vocal cords, whereas voiceless consonants -- such as /p, t, k, f, s/ -- do not involve such vibrations. There is no doubt that speakers of English hear the consonants within each of these two sets as similar. An analysis of the confusions resulting when subjects listen to consonant-vowel syllables embedded in white noise and identify them (for example, Miller & Nicely, 1955) reveals that confusions occur largely within voicing category. In fact, not only do the perceptual confusions of consonants reflect the voicing distinction, but also confusion errors in short-term memory reflect the voicing distinction, even when the to-be-remembered letters are presented visually (see, for example, Wickelgren, 1966; Hintzman, 1967). Furthermore, estimates of the similarity between various

---

*This paper will appear in Memory & Cognition.

+Also Yale University.

++Also Wellesley College.

Acknowledgment. This research was supported in part by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories and NSF Grant BNS77-00077 to Yale University. The authors are deeply grateful to Rick Arons for his help with the design, conduct, and analyses of Experiment 3 and his helpful comments on an earlier version of this manuscript; Loretta Polka for her help with the design, conduct, and analyses of Experiments 1 and 2; Bruno Repp and Guy Carden for their helpful comments about this research, and Robert Crowder for his useful comments on an earlier version of this manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

consonant sounds compared in pairs are relatively large when the sounds agree in voicing (Greenberg & Jenkins, 1964). These results demonstrate that when considered in pairs, consonants in the same voicing category are deemed similar; however, the question remains whether subjects can abstract the voicing distinction across such pairs. In other words, despite the clear effects on behavior of the voicing distinction, the voicing distinction per se may not be accessible for use in learning and memory tasks (cf. LaRiviere, Winitz, Reeds, & Herriman, 1974). Do adult speakers of English without formal training in linguistics know that /b, d, g, v, z/ fall within one class of consonants, whereas /p, t, k, f, s/ fall within another, even if they do not know the labels for these classes?

At first glance, the answer to this question seems obvious since numerous phonological rules make reference to the voicing distinction. For example, one aspect of the rule for forming the plural of English nouns[1] involves voicing assimilation: The voicing of the final sibilant /s/ or /z/ representing the plural morpheme must be the same as that of the immediately preceding segment for stems ending in /p, t, k, f, b, d, g, v/. Thus, the sound of the plural morpheme will be /s/ for stems ending in /p, t, k, f/ and will be /z/ for stems ending in /b, d, g, v/. One might argue that if individuals know this rule, they must know the distinction between voiced and voiceless consonants. Such an argument seems implied by the "feature" hypothesis of English pluralization discussed by Derwing and Baker (1977). On the basis of this line of reasoning, one would answer the question about the accessibility of the voicing distinction affirmatively, not only for adults but also for children 4-7 years old, since even young children have demonstrated the ability to use this plural formation rule productively (Berko, 1958). However, it could also be argued that use of this phonological rule does not require knowledge of the voicing distinction because the rule follows from phonetic or articulatory constraints on the production of consonant clusters[2] (see, for example, the "phonotactic" rules discussed by Derwing & Baker, 1977). It is impossible for an individual to pronounce /z/ following /p, t, k, f/ in a single syllable. For that reason, whether the voicing distinction is accessible is not only an open question, but also a very interesting one, since it should illuminate how individuals learn and use the phonological rules involving voicing assimilation.

The aim of the present study was to determine how accessible the distinctive feature <u>voicing</u> is to adult speakers of English. Two paradigms developed by Healy and Levitt (1978) to study the accessibility of syntactic constructs were used in the present investigation to explore the accessibility of the voicing distinction. The first paradigm involved a recognition memory task that allowed us to determine how readily subjects could discover a rule based on the voicing distinction in their attempts to learn a list of nonsense syllables for a subsequent memory test. The second paradigm involved a more conventional concept formation task that allowed us to determine the extent to which subjects could learn (by example) a rule based on voicing. In both paradigms, the rule based on the voicing distinction was compared to an arbitrary rule based on a division of the consonants into two sets, each of which contained both voiced and voiceless members -- <u>b</u>, <u>d</u>, <u>k</u>, <u>f</u> in one set and

p, t, g, v in the other set. Also in both paradigms, nonsense syllables ending in consonants were employed as stimuli, and both rules, like the plural formation rule, made reference to the final consonant in the syllables: Syllables ending in one set of consonants were followed by one terminal symbol (for example, s), and syllables ending in the other set of consonants were followed by a second terminal symbol (for example, z). The terminal symbols employed were varied so that they were either distinguishable on the basis of voicing (as in s and z) or were not distinguishable on the basis of voicing (as in ! and ?).

EXPERIMENT 1

The recognition memory paradigm was employed in Experiment 1. In this experiment subjects studied a series of nonsense syllables presented visually. In the "sz" condition, half of these syllables were followed by the letter s and the other half by the letter z; in the "!?" condition, the symbols ! or ? were used instead. At the time of test, the subjects were shown the syllables without their terminal symbols and were asked to indicate for each syllable whether it had been terminated by s or z (or ! or ?) at the time of study. The rule determining the assignment of the terminal symbol was based either on the voicing of the final consonants (the "Voicing" conditions) or on the arbitrary division of the final consonants (the "Arbitrary" conditions). If the voicing distinction is accessible, subjects in the Voicing conditions should perform better than those in the Arbitrary conditions. If, on the other hand, subjects can easily discover a rule based on voicing only if it involves voicing assimilation of final consonants, then subjects in the Voicing sz condition should perform better than subjects in each of the other conditions.

Method

Subjects. Sixty young men and women participated in this experiment. The first 20 subjects were Yale undergraduates who participated for course credit. The remaining 40 subjects were individuals who responded to advertisements posted on the Yale University campus and were paid $1.25 for their participation. All subjects were native speakers of English who had had no formal training in linguistics. The subjects were divided into four groups of 15: Voicing sz, Voicing !?, Arbitrary sz, Arbitrary !?. The assignment of subjects to the four groups was determined by time of arrival for testing according to a fixed rotation of conditions.

Design and materials. Sixty-four different nonsense syllables were employed as stimuli. The nonsense syllables ended in one of eight consonants -- b, d, g, p, t, k, f, v -- and began with one of eight vowel pairs -- ae, ai, au, oe, oi, ou, aa, oo -- (e.g., aeb, aid). Each of the 64 syllables followed by a terminal character (see below) was typed in the center of four 3 x 5 in. cards. Four decks of cards were constructed for use during the study phase of the experiment, one for each of the four conditions. Each deck included all 64 syllables. The terminal characters for two of the four decks

(Voicing sz and Arbitrary sz) were s and z and for the other two decks (Voicing !? and Arbitrary !?) were ! and ?. In the decks used for the Voicing conditions, the characters s or ! followed all syllables ending in the voiceless consonants p, t, k, or f, and the characters z or ? followed all syllables ending in the voiced consonants b, d, g, or v. In the decks used for the Arbitrary conditions, the characters s or ! followed all syllables ending in the consonants p, t, g, or v, and the characters z or ? followed all syllables ending in the consonants b, d, k, or f. The experimenter thoroughly shuffled the deck of cards before handing it to a given subject, so that the order of the syllables in a given deck varied across subjects.

Four typewritten lists of syllables were constructed for the recognition memory test, one list for each of the four conditions. Each syllable on the two lists used in the sz conditions was followed by the pair of responses written as (s/z), whereas each syllable on the two lists used in the !? conditions was followed by (!/?). Each of the four lists included all 64 syllables; only the order of the syllables varied across lists. In each list, the order of syllables was pseudo-random with the constraints that each 16-syllable block included two syllables with each consonant and that the order of correct answers (s or z, or ! or ?) was the same for subjects in the Voicing and Arbitrary conditions.

Procedure. Subjects were tested individually in a single session that lasted approximately 15 minutes. Each subject was given five minutes to study one of the four decks of cards (timed by the experimenter with a stopclock), and was warned when the first 2.5 minutes had elapsed. Subjects were in no way restricted in their method of studying the syllables, except that they were told to pronounce each syllable aloud as a single syllable when they read it. They were allowed to sort the syllables into piles, and they were allowed to look at a given syllable any number of times. The subjects were not encouraged to use any particular strategy in studying the syllables. They were, however, given written instructions describing exactly what their task would be during the recognition memory test. The instructions for the sz conditions were the following: "You will be presented with a stack of cards. On each card is a nonsense syllable which ends in either s or z. You are to study these nonsense syllables for five minutes. Pronounce each nonsense syllable aloud as a single syllable when you read it; say it loud enough so that the experimenter can hear you. At the end of five minutes you will be given a sheet of paper which includes each of the nonsense syllables on the cards with the final letter s or z replaced by (s/z). Your task will be to recall for each nonsense syllable whether s or z was at the end of that syllable when it appeared on the card. You are to indicate your response by circling one of the two letters s or z at the end of a given nonsense syllable on the sheet of paper. Before you respond to a given syllable, you are to pronounce it aloud." The instructions for the !? conditions were identical except that the letters s and z were replaced by the symbols ! and ?, respectively.

After studying the syllables on the cards, the subjects were reminded of their task on the recognition memory test. The subjects were then given the

appropriate test list of syllables. They responded to each syllable by pronouncing it aloud and then circling one of the two terminal symbols (s or z, or ! or ?), depending on which symbol they thought occurred with the syllable when it appeared on the card. Subjects were required to respond to every nonsense syllable; they were not allowed to leave blanks. Subjects were given as much time as they needed to complete the recognition memory test.

Results

The results are summarized in Table 1 in terms of mean percentages of errors on the recognition test as a function of condition. Subjects in the Voicing sz group made fewer errors than subjects in the remaining three groups. An analysis of variance performed on these data revealed a significant effect of rule type (voicing or arbitrary), [$F(1,56) = 16.9$, $MSe = 251.2$, $p < .001$], as well as a significant effect of terminal symbols (sz or !?), [$F(1,56) = 5.5$, $MSe = 251.2$, $p = .021$], but the interaction of these two factors was not significant, [$F(1,56) = 1.3$, $MSe = 251.2$, $p = .251$].

Table 1

Mean Percentage of Errors in Experiment 1
as a Function of Condition

	Terminal symbols	
Rule	sz	!?
Voicing	10.0	24.4
Arbitrary	31.6	36.5

Planned analyses of variance conducted on each pair of terminal symbols separately yielded a significant effect of rule type for the sz pair of endings [$F(1,28) = 18.9$, $MSe = 184.2$, $p < .001$], but not for the !? pair of endings [$F(1,28) = 3.4$, $MSe = 318.3$, $p = .071$]. Although the effect of rule type was not significant for !? by the standard two-tailed test, it was by a one-tailed test, so the results would appear somewhat ambiguous for !? from this analysis alone. However, two further analyses indicate that !? does in fact show a smaller effect of rule type than sz. First, planned analyses conducted on each rule type separately yielded a significant effect of terminal symbols for the voicing rule [$F(1,28) = 6.7$, $MSe = 230.6$, $p = .014$], but not for the arbitrary rule [$F(1,28) < 1$]. Second, an analysis of the number of subjects who made no errors showed that six subjects made no errors in the Voicing sz group, but only one subject made no errors in the Arbitrary sz group. In contrast, for both the Voicing !? and the Arbitrary !? groups, two subjects made no errors.

Discussion

A large advantage was found for the Voicing over the Arbitrary conditions for the sz pair of endings. In contrast, a smaller difference was found between performance levels in the Voicing and Arbitrary conditions for the !? pair of endings. These findings suggest that subjects can discover a rule based on voicing more easily if the rule involves voicing assimilation than if no assimilation is involved. The most straightforward explanation of these results seems to be based on the fact that voicing assimilation follows from articulatory constraints: Subjects know, for example, that ps is an acceptable final consonant cluster but not pz, simply because they are unable to pronounce /pz/ at the end of a single syllable. However, an alternative explanation for the results of this experiment is available. Possibly, subjects perform better in the Voicing sz condition than in the other three conditions because of the overt similarity to the plural formation rule in English. Subjects can determine the correct answers in the Voicing sz condition merely by treating the terminal symbol as the plural morpheme and equating the letter s with the phoneme /s/ and the letter z with the phoneme /z/.

EXPERIMENT 2

In order to distinguish between the two explanations proposed above for the results of Experiment 1, a new pair of terminal symbols was selected for examination in Experiment 2 -- f and v. These two letters, like s and z, are distinguishable on the basis of voicing, but unlike s and z, their relationship to the sounds representing the plural morpheme in English is not so straightforward and, in fact, they do not occur following stop consonants in final consonant clusters in English. In addition, s and z were employed in this experiment for comparison, as well as a third pair of symbols -- m and n. The third pair was selected to replace the symbols ! and ? used in Experiment 1, because like ! and ?, m and n are not distinguishable on the basis of voicing, but unlike ! and ?, m and n are letters and thus permit a comparison of symbol pairs under conditions as analogous as possible. In addition to this change in terminal symbol pairs, Experiment 2 involved a change in paradigm. A concept formation task, rather than the recognition memory procedure, was employed, in order to enable us to determine the generality of the findings in Experiment 1. The concept formation procedure is a more standard experimental technique and has been used by other investigators to study phonological rules.3

In this experiment, as in Experiment 1, a series of nonsense syllables was presented visually to subjects. Unlike Experiment 1, only stop consonants were employed in the syllables, since f and v were used as terminal letters in some conditions. The syllables were shown successively in a fixed order. For each syllable, the subjects had to choose the appropriate terminal letter, which was assigned according to the rule based on voicing of the final consonants or the rule based on the arbitrary grouping of final consonants. Subjects were provided immediate feedback after responding to each syllable.

Half the subjects in each group were asked to pronounce the syllables aloud with their endings both at the time of responding and after feedback was provided, and the other half were given no explicit instructions to pronounce the syllables aloud. Subjects performed this task in the guise of learning a rule for gender formation in an artificial language, in which one of a pair of terminal letters was used with masculine words and the other with feminine words.

If the voicing distinction is accessible under these conditions, subjects learning the voicing rule should perform better than those learning the arbitrary rule. If, on the other hand, subjects can easily learn a rule based on voicing only if it involves voicing assimilation, the subjects learning the voicing rule with the terminal letter pairs sz and fv should perform better than subjects in each of the other groups. However, if subjects can easily learn a rule involving voicing assimilation only when they are attending to the sound and articulation of the syllables, then subjects in the Voicing sz and fv conditions who are asked to pronounce the syllables aloud should perform better than those who are not so instructed. Alternatively, if subjects can easily learn a voicing rule only if its relationship to the plural formation rule in English is straightforward, then subjects learning the voicing rule with the terminal letter pair sz should perform better than subjects in all the remaining conditions.

Method

Subjects. Forty-eight male and female Yale undergraduates participated for course credit. All subjects were native speakers of English who had had no formal training in linguistics. The subjects were divided into six groups of eight: Voicing sz, Voicing fv, Voicing mn, Arbitrary sz, Arbitrary fv, Arbitrary mn. Each group of eight subjects was further subdivided into two subgroups of four: Aloud and Silent. The assignment of subjects to groups and subgroups was determined by time of arrival for testing according to a fixed rotation of groups and subgroups.

Apparatus. An Adds 980 terminal, including a typewriter keyboard and a CRT screen and controlled by a PDP-11/45 computer operating under a time-sharing system, was used to display the stimuli and receive the subjects' responses.

Design and materials. Sixty different nonsense syllables were employed as stimuli, which were similar to those used in Experiment 1 except that six consonants were employed -- b, d, g, p, t, k -- and ten vowel pairs -- ae, ai, au, oe, oi, ou, aa, oo, eu, ie. Two lists of syllables were constructed, one for the Voicing conditions and one for the Arbitrary conditions. Each list included all 60 syllables; only the order varied across lists. In each list the order of the syllables was pseudo-random with the constraints that each 12-syllable block included two syllables with each consonant and that the sequence of correct responses (the correct terminal letters) was the same in the Voicing and Arbitrary conditions.

Procedure. Each subject was tested individually in a single session lasting approximately 15 minutes. In each condition, the subject saw at the center-top of the display screen the 60 syllables from the appropriate list, one at a time in the prescribed order. The subject was to respond to each syllable by typing at the end of the syllable one of the two terminal letters for the condition, depending on whether the syllable was "masculine" or "feminine" (see below). After the subject responded, the computer supplied immediate feedback below the display of the syllable and the response letter in the form of the statement, "Correct/Wrong, the answer is: -- ," where the blank was filled by the syllable with its appropriate terminal letter. When the subject was ready for the next trial, he or she was then to press the key "new line" and the screen was cleared and the next syllable appeared on the screen. Syllable presentation was thus subject paced. In the sz conditions, half the syllables were followed by s and half by z; in the fv conditions, the syllables were followed by either f or v; and in the mn conditions, the syllables were followed by either m or n. In the Voicing conditions all syllables ending in one of the voiceless consonants -- p, t, or k -- were matched with one of the terminal letters s, f, or m, and all syllables ending in one of the voiced consonants -- b, d, or g -- were matched with one of the corresponding terminal letters z, v, or n. In the Arbitrary conditions the syllables were divided into those ending in p, t, or g and those ending in b, d, or k.

Subjects in the Aloud subconditions were explicitly told to pronounce aloud each syllable with its ending as a complete word both at the time of responding and after feedback was given, whereas subjects in the Silent subconditions received no such instructions.

The subjects were instructed that the nonsense syllables they would see represented root words in an artificial language and that gender in the language was denoted by word endings. Half the subjects in each subcondition were told that in the artificial language masculine words ended in s (or f or m, depending on the condition) and feminine words ended in z (or v or n, depending on the condition). For the remaining half of the subjects, the instructions concerning gender were reversed, so that feminine words were said to end in s, f, or m, and masculine words in z, v, or n. This manipulation insured that the terminal letter was not confounded with gender. The subjects were asked to determine whether each syllable was masculine or feminine, and, thus, which member of the terminal letter pair was appropriate for the syllable.

Subjects were told that in the beginning they would have to guess which syllables were masculine and which were feminine, but as the session progressed, they should be able to learn by example the rule that determined gender in the artificial language.

Results

The results are summarized in Table 2 in terms of mean percentages of errors on the concept formation task as a function of condition and

subcondition. Although an advantage for the voicing rule over the arbitrary rule was found for the sz pair of endings, no difference between performance on the two rules was found for the fv or mn ending pairs. In fact, there was a small advantage for the arbitrary rule over the voicing rule for the fv pair of endings. Neither the main effect of rule type (voicing or arbitrary) [$F(1,36) = 2.0$, $MSe = 1066.7$, $p = .162$] nor the main effect of terminal letter pair (sz or fv or mn) [$F(2,36) < 1$] were significant, but the interaction of the two factors was significant [$F(2,36) = 3.4$, $MSe = 1066.7$, $p = .044$]. Planned analyses of variance conducted on each pair of endings separately yielded a significant effect of rule type for the sz pair [$F(1,12) = 9.1$, $MSe = 939.0$, $p = .011$] but not for either the fv pair [$F(1,12) < 1$] or the mn pair [$F(1,12) < 1$].

Table 2

Mean Percentage of Errors in Experiment 2
as a Function of Condition and Subcondition

	Terminal letter pair					
	sz		fv		mn	
Subcondition	Voicing	Arbitrary	Voicing	Arbitrary	Voicing	Arbitrary
Aloud	21.3	49.2	39.6	35.4	40.0	35.8
Silent	29.6	42.9	44.6	37.1	37.1	47.5
Mean	25.4	46.0	42.1	36.2	38.5	41.7

The factor of subcondition (Aloud vs. Silent) was not significant as a main effect [$F(1,36) < 1$] and did not enter into any significant interactions.

Learning was evidenced by improvement in performance across the five twelve-trial blocks. The main effect of blocks was significant [$F(4,144) = 12.9$, $MSe = 205.4$, $p < .001$]. However no differences in the extent of learning as a function of condition were evident. None of the interactions involving the factor of blocks were significant. The differences among conditions therefore emerged within the first block of 12 trials.

Discussion

No difference was found between levels of performance in the Voicing and Arbitrary conditions for the terminal letter pairs mn and fv. This finding, in accord with that for !? in Experiment 1, provides further support for the hypothesis that the voicing distinction is not accessible for use in memory

and learning tasks. In contrast, but in agreement with Experiment 1, an advantage was found for the Voicing over the Arbitrary conditions with the sz pair of endings. Because of the different pattern of results for the sz and fv pairs, these results do not support the hypothesis that subjects can easily learn a rule based on voicing only if it involves voicing assimilation, since voicing assimilation applies for fv as well as sz. Instead, these results are consistent with the hypothesis that subjects can easily learn a voicing rule only if its relationship to the plural formation rule in English is straightforward.

Perhaps subjects did not easily learn the rule based on voicing assimilation with the fv endings because f and v do not follow stop consonants in final consonant clusters in English. For this reason subjects working with the fv endings might have resisted attending to the sound and/or articulation of the syllables. Furthermore, although there were no significant differences between the Aloud and Silent subconditions, subjects in the Voicing fv and Voicing sz conditions both made fewer errors in the Aloud than in the Silent subconditions. In order to examine more thoroughly the importance of sound and articulation, we turned to the auditory, rather than the visual, modality in Experiment 3. The use of the auditory modality should encourage attention to the sound and articulation of the syllables rather than their visual features and therefore might lead to better performance with the fv endings. Also, the auditory modality is of interest in this context, since it is the modality used by children when learning a language.

EXPERIMENT 3

Experiment 3 was essentially a replication of Experiment 2 except for two changes: (1) the stimuli were presented auditorily rather than visually, and (2) the terminal letters a and o, which are also not distinguishable on the basis of voicing, were used instead of m and n. The letter pairs were changed in order to make it easier for the subjects to pronounce the syllables with their endings. The initial aural presentation of the stimuli without the endings in this experiment allowed us to make sure that the subjects would clearly hear the distinction between the voiced and voiceless consonants.

If the change to the auditory modality does draw attention from the irrelevant visual properties of the syllables, thereby forcing subjects to rely on the sound and articulation of the syllables, the subjects learning the voicing rule with the fv endings should perform better than those learning the arbitrary rule with the fv endings. If, on the other hand, the change to the auditory modality does not have the intended effects, the results of this study should essentially replicate those of Experiment 2.

Once again subjects in each group were divided into two subgroups -- those required to pronounce aloud the syllables with their endings (Aloud) and those not given any explicit instructions to do so (Silent). We expect to find an advantage for the Aloud subgroup in the Voicing sz and Voicing fv conditions, if subjects can learn a rule based on voicing assimilation only when they are attending to the sound and articulation of the syllables.

Method

Subjects. Seventy-two male and female Yale undergraduates were employed; 65 received course credit and the remaining 7 participated purely on a volunteer basis. All subjects were native speakers of English who had had no formal training in linguistics. As in Experiment 2, the subjects were divided into six equal groups. The groups were analogous to those used in Experiment 2 except that the mn groups were replaced by ao groups. Also as in Experiment 2, each group of subjects was further subdivided into Aloud and Silent subgroups. The assignment of subjects to groups and subgroups was determined by time of arrival for testing according to a fixed rotation of groups and subgroups.

Apparatus. The stimuli (syllables, names of the terminal letters, and the phrase "the correct answer is") were recorded on an Ampex AG 500 tape recorder in a soundproof room by a female native speaker of English (AGL), who clearly articulated the syllables and released the final stop consonants so that the distinction between the voiced and voiceless consonants was obvious. These stimuli were digitized using the Haskins Laboratories Pulse Code Modulation System (Cooper & Mattingly, 1969). Subsequently, the stimuli were edited for starting point and end point, by adjusting the duration of the silent periods at the start and end, to insure that the presentation times for all syllables were equal (640 msec), as were those for the names of all the terminal letters (660 msec). The stimuli were reconverted into analog form and recorded on both channels of a two-channel Crown 800 tape recorder. The use of the Pulse Code Modulation System insured that all instances of a given stimulus on the tapes were identical.

The stimuli were transmitted to the subject binaurally through a pair of Telephonics earphones model TDH-39. The stimulus tapes were played with a TEAC A-3300S tape recorder at a comfortable listening level.

Design and materials. Sixty different nonsense syllables were employed as stimuli, which were analogous to those used in Experiment 2. As in Experiment 2, six consonants were employed -- /b, d, g, p, t, k/ -- and ten vowels -- /i, I, e, ɔ, ʌ, u, o, ɛ, æ, a/. Six tapes were constructed, one for each of the six conditions. Each tape included 60 trials. A trial consisted of the following sequence of events: (1) the presentation of a vowel-consonant syllable, (2) a 6-sec silent interval, (3) the statement, "the correct answer is," (4) a 590-msec silent interval, (5) the name of the correct terminal letter, (6) a 3-sec silent intertrial interval. At the end of every block of 12 syllables was a 7-sec silent interval in addition to the 3-sec silent intertrial interval. The order of presentation of the syllables on the tapes corresponded exactly to the order used in Experiment 2 with the vowels /æ, e, ɔ, ʌ, i, ɛ, I, u, a, o/ replacing the vowel pairs ae, ai, au, eu, ie, oe, oi, ou, aa, oo, respectively. The tapes for the sz, fv, and ao conditions were identical except that the terminal letters s and z on the sz tapes were replaced by f and v on the fv tapes and a and o on the ao tapes.

Procedure. As in Experiment 2, each subject was tested individually in a single session lasting approximately 25 minutes. Each subject participated in one of the six conditions, defined by the rule to be learned -- voicing or arbitrary -- and the terminal letter pair -- sz, fv, or ao -- and one of two subgroups -- Aloud or Silent. The subjects were to respond to each nonsense syllable by circling one of the two terminal letters beside the trial number on an answer sheet. The subjects were given six seconds to make each response. As in Experiment 2, subjects in the Aloud subconditions, but not those in the Silent subconditions, were told to pronounce aloud each syllable with its ending as a complete word both at the time of responding and after feedback was given in the form of the name of the correct terminal letter.

The instructions given to the subjects were analogous to those used in Experiment 2, except that the subjects were warned to be prompt in making their responses because the timing was fixed in advance and could not be changed.

Results

The results are summarized in Table 3 in terms of mean percentages of errors on the concept formation task as a function of condition and subcondition. An advantage for the voicing rule over the arbitrary rule was found for the sz and the fv pairs of endings, but no difference between the two rules was found for the ao pair. The main effect of rule type (voicing or arbitrary) was significant [$F(1,60) = 8.7$, $MSe = 892.8$, $p = .005$], but the main effect of terminal letter pair (sz or fv or ao) [$F(2,60) = 2.3$, $MSe = 892.8$, $p = .103$], and the interaction of rule type and terminal letter pair [$F(2,60) = 1.6$, $MSe = 892.8$, $p = .211$] were not significant. Planned analyses of variance conducted on each pair of endings separately yielded a significant effect of rule type for the sz pair [$F(1,20) = 6.6$, $MSe = 863.1$, $p = .018$], and for the fv pair [$F(1,20) = 5.1$, $MSe = 954.9$, $p = .033$], but not for the ao pair [$F(1,20) < 1$]. Additional planned analyses conducted on each rule type separately yielded an effect of terminal symbol which just missed significance for the voicing rule [$F(2,30) = 3.0$, $MSe = 1056.1$, $p = .062$], but did not approach significance for the arbitrary rule [$F(2,30) < 1$].

The subcondition manipulation did prove to be important in this experiment. Specifically, subjects in the Voicing sz and Voicing fv conditions performed somewhat better in the Aloud subconditions than in the Silent subconditions, but a difference in the opposite direction was found for subjects in the four remaining conditions. The three-way interaction of terminal letter pair, subcondition, and rule type was not significant in the overall analysis [$F(2,60) = 1.2$, $MSe = 892.8$, $p = .309$]; however, the two-way interaction of terminal letter pair and subcondition was significant in the overall analysis [$F(2,60) = 3.5$, $MSe = 892.8$, $p = .036$]. Further, the analyses conducted on the two rule types separately revealed a significant interaction of terminal letter pair and subcondition for the voicing rule [$F(2,30) = 3.6$, $MSe = 1056.1$, $p = .039$], but not for the arbitrary rule [$F(2,30) < 1$].

Table 3

Mean Percentage of Errors in Experiment 3
as a Function of Condition and Subcondition

	Terminal letter pair					
	sz		fy		ao	
Subcondition	Voicing	Arbitrary	Voicing	Arbitrary	Voicing	Arbitrary
Aloud	18.9	42.2	24.4	42.5	45.3	44.4
Silent	37.8	41.9	25.3	32.8	32.5	36.1
Mean	28.3	42.1	24.9	37.6	38.9	40.3

Table 4

Mean Percentage of Errors in Experiment 3
as a Function of Rule Type, Subcondition, and Final Consonant

	Rule Type			
	Voicing		Arbitrary	
Final consonant	Aloud	Silent	Aloud	Silent
B	25.0	33.3	47.2	42.2
D	28.3	32.8	33.3	28.3
G	30.6	30.0	51.7	33.3
P	33.3	23.3	35.0	40.0
T	32.8	32.2	35.0	37.2
K	27.2	39.4	56.1	40.6

An examination of the pattern of errors as a function of final consonant suggests that subjects may have had a natural tendency to use the voicing rule in the Arbitrary Aloud subcondition: The greatest number of errors in that subcondition occurred for syllables ending in g or k, the only two of the six final consonants that violated the voicing rule. This pattern of errors is depicted in Table 4, which provides mean error percentages as a function of rule type, subcondition, and final consonant. In the overall analysis of variance, the main effect of final consonant was significant [$F(5,300) = 3.6$, $MSe = 252.0$, $p = .004$], as was the interaction of rule type and final consonant [$F(5,300) = 2.8$, $MSe = 252.0$, $p = .018$], and the second-order interaction of rule type, subcondition, and final consonant [$F(5,300) = 4.1$, $MSe = 252.0$, $p = .001$]. These results illustrate the salience of the voicing rule in the Aloud subconditions.

Learning was evident across the five twelve-trial blocks [$F(4,240) = 15.7$, $MSe = 173.3$, $p < .001$]. Although none of the interactions involving blocks were significant in the overall analysis, the interaction of terminal letter pair and blocks was significant in the analysis of the voicing rule [$F(8,120) = 2.2$, $MSe = 179.4$, $p = .030$], reflecting the greater learning of the voicing rule in the sz and fv conditions than in the ao condition. This interaction is depicted in Figure 1, which provides mean error percentages as a function of trial block position and terminal letter pair for the voicing rule. The corresponding interaction was not significant for the arbitrary rule [$F(8,120) < 1$], although the main effect of blocks was significant for both the arbitrary rule [$F(4,120] = 7.6$, $MSe = 167.4$, $p < .001$] and the voicing rule [$F(4,120) = 8.5$, $MSe = 179.4$, $p < .001$], reflecting the fact that some learning took place for both rules. In the analysis conducted on each of the terminal letter pairs separately, including both voicing and arbitrary rules, the main effect of blocks was significant for the sz pair [$F(4,80) = 5.3$, $MSe = 177.3$, $p = .001$] and the fv pair [$F(4,80) = 12.4$, $MSe = 162.6$, $p < .001$], but not for the ao pair [$F(4,80) = 1.9$, $MSe = 180.1$, $p = .122$], providing further support for greater learning in the sz and fv conditions than in the ao condition.

Discussion

No difference was found between performance on learning the voicing and arbitrary rules for the terminal letter pair ao. This finding is consistent with the findings for the terminal symbol pairs !? in Experiment 1 and mn in Experiment 2. Since the distinction between the final voiced and voiceless stop consonants was made clear by the initial aural presentation of the syllables, this result gives support to the hypothesis that the voicing distinction is not accessible for use in learning tasks. As in Experiments 1 and 2, an advantage was found for the Voicing over the Arbitrary conditions with the sz pair of endings. Furthermore, unlike Experiment 2, an advantage was also found in this experiment for the Voicing over the Arbitrary conditions with the fv pair of endings. This finding suggests that use of the auditory modality does in fact draw attention away from the irrelevant visual properties toward the sound and articulation of the syllables, and that under such conditions subjects can easily learn a rule involving voicing

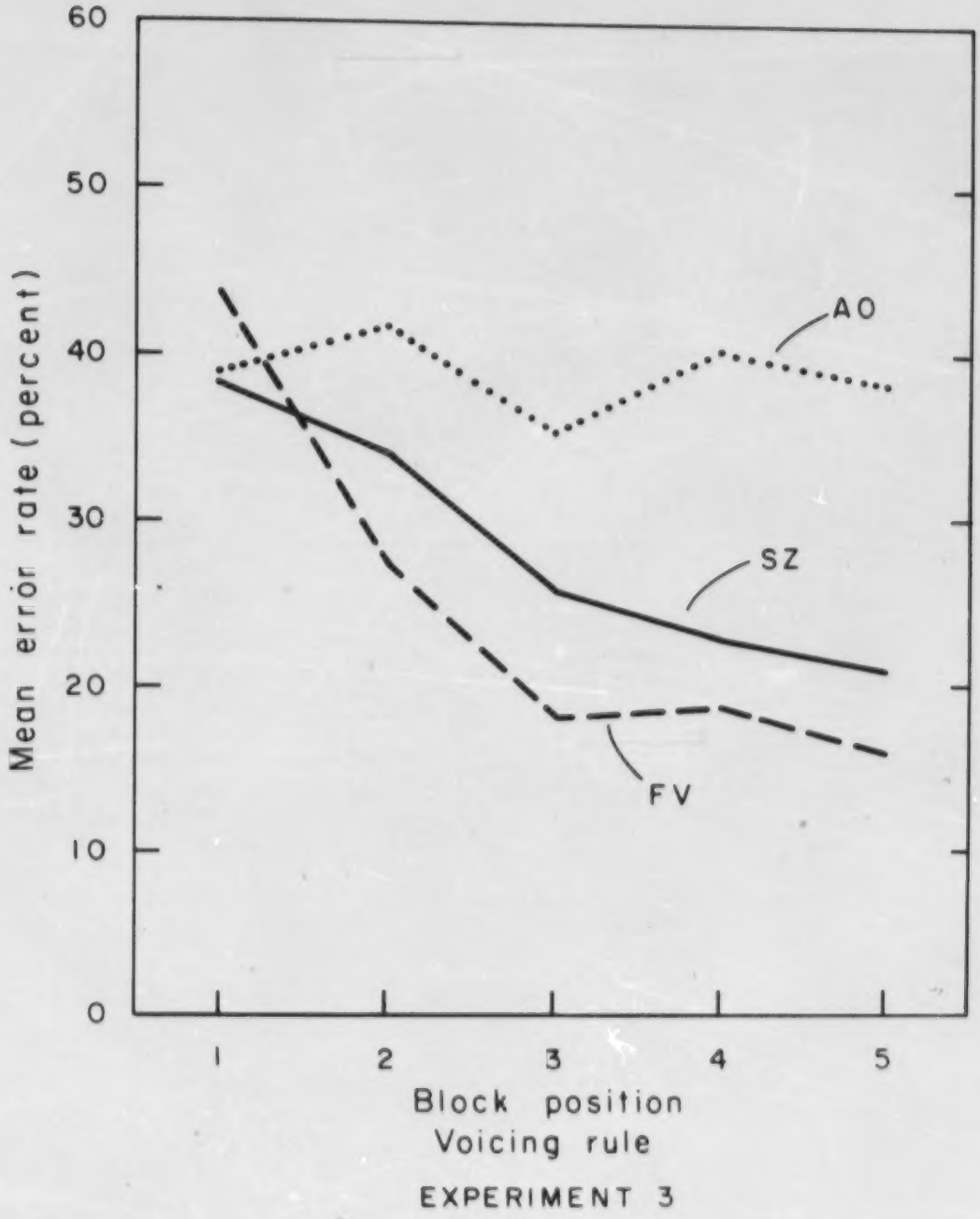


Figure 1. Mean percentage of errors in the Voicing conditions of Experiment 3 as a function of terminal letter pair and block position.

assimilation. The importance of articulation of the syllables for learning the voicing assimilation rule was further supported by the fact that there was some advantage for the Aloud over the Silent subconditions in the Voicing conditions when the terminal letter pair was sz or fv (but not when it was ao and not in the Arbitrary conditions).

A paired associate learning experiment by Jenkins, Foss, and Greenberg (1968) also demonstrated the importance of articulation in learning a rule based on distinctive features. Subjects in this experiment learned a list of six pairs of consonant-vowel syllables. For the control group of subjects the response terms were randomly paired with the stimulus terms. For the experimental groups of subjects, the stimuli and responses were identical in terms of all features except voicing (for example, pa was paired with ba). The several experimental groups differed in the instructions given to them directing their attention to the type of relationship between the stimuli in a pair. For example, one experimental group of subjects was told to attend to "what your mouth is doing as you say the syllables" (p. 202). This group performed better than the other experimental groups, who were not directed to the articulation of the syllables, and these groups in turn performed better than the control group.

SUMMARY AND CONCLUSIONS

We have found that subjects can discover or learn a rule based on the distinctive feature voicing more easily than a rule based on an arbitrary division of consonants only when the relationship between the voicing rule and the plural formation rule is straightforward or, more generally, when the voicing rule involves voicing assimilation and attention is paid to the sound and articulation of the stimuli. These results suggest that the voicing distinction is not ordinarily accessible to subjects for use in memory and learning tasks. This conclusion should not be taken to imply that the voicing distinction does not play some role in perception and memory. In fact, to the contrary, as reviewed in the introduction, there is considerable evidence that the voicing distinction does have important effects on memory and perception. However, the conclusion reached here concerning the inaccessibility of the voicing distinction is inconsistent with the claim that learning and use of various phonological rules in English such as the plural formation rule depend on the analysis of voicing as a distinctive feature (cf. the "feature" hypothesis of Derwing & Baker, 1977). We propose instead that individuals easily learn and use phonological rules involving voicing assimilation because of articulatory constraints. Subjects know that syllables ending in /p, t, k/ are followed by /s/ and syllables ending in /b, d, g/ are followed by /z/ simply because they are unable to pronounce syllables ending in consonant clusters /pz, tz, kz/. This study vividly illustrates that the form in which a linguist most clearly specifies a given rule need not be the form in which the rule is learned and used by speakers of the language.

The fact that the voicing feature is not ordinarily accessible for use in learning and memory does not necessarily imply that other distinctive features are also inaccessible. In fact, in a categorization paradigm, in which each

of a small set of consonant-vowel syllables differing only in their initial consonants was placed into one of two categories, LaRiviere et al. (1974) found that the voicing distinctive feature was not accessible but other distinctive features (for example, the strident feature) were. It would be of interest to determine whether such a difference in the accessibility of features is also found with the procedures developed in the present study.

REFERENCE NOTES

1. Cena, R. M. When is a phonological generalization psychologically real? Indiana University Linguistics Club, 1978.
2. Dinnsen, D. A. Phonological rules & phonetic explanation. Indiana University Linguistics Club, 1978.

REFERENCES

Berko, J. The child's learning of English morphology. Word, 1958, 14, 150-177.

Cooper, F. S., & Mattingly, I. G. Computer-controlled PCM system for investigation of dichotic speech perception. Haskins Laboratories Status Report on Speech Research, 1969, SR-17/18, 17-21.

Derwing, B. L., & Baker, W. J. The psychological basis for morphological rules. In J. Macnamara (Ed.), Language learning and thought. New York: Academic Press, 1977.

Greenberg, J. H., & Jenkins, J. J. Studies in the psychological correlates of the sound system of American English. Word, 1964, 20, 157-177.

Healy, A. F., & Levitt, A. G. The relative accessibility of semantic and deep-structure syntactic concepts. Memory & Cognition, 1978, 6, 518-526.

Hintzman, D. L. Articulatory coding in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 312-316.

Jakobson, R., Fant, C. G. M., & Halle, M. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: M.I.T. Press, 1963.

Jenkins, J. J., Foss, D. J., & Greenberg, J. H. Phonological distinctive features as cues in learning. Journal of Experimental Psychology, 1968, 77, 200-205.

LaRiviere, C., Winitz, H., Reeds, J., & Herriman, E. The conceptual reality of selected distinctive features. Journal of Speech and Hearing Research, 1974, 17, 122-133.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustical Society of America, 1966, 39, 388-398.

FOOTNOTES

[1]The plural formation rule in English is phonetically equivalent to the rule for forming possessives of English nouns and the third-person singular of

the present tense of English verbs. We shall refer to all these rules collectively as the "plural formation rule."

[2]Progressive voice assimilation, in which the second phoneme in a consonant cluster is made to agree in voicing with the first phoneme, is the solution used in English to the articulatory constraints. Dinnsen (Note 2) points out that other languages employ another solution to the articulatory constraints -- regressive voice assimilation, in which the first phoneme is made to agree in voicing with the second phoneme.

[3]Cena (Note 1) reviewed a set of studies on vowel alternation rules and discussed the relative merits of the experimental techniques employed in those studies. He concluded that the concept formation technique was preferable to some of the others employed because of its built-in control for response bias, a control also incorporated in the recognition memory procedure used in Experiment 1.

fricative noise leaves the perceptual effect unchanged, whereas insertion of a silent interval following the noise leads to a substantial reduction of the contextual dependency. We further tried to determine whether it is mere temporal separation or the perception of an intervening stop consonant (caused by the silent interval inserted) that is responsible for this reduction. The results suggested temporal separation as the primary factor, which agrees with recent, analogous observations on anticipatory lip rounding (Bell-Berti & Harris, 1979).

In the present paper, we continue our investigations of the effect of vocalic context on fricative perception (often referred to in the following as the "vowel context effect"). Whereas, in the earlier paper (Mann & Repp, 1979a), we were concerned primarily with temporal factors (duration of the fricative noise and duration of a silent interval following the noise), the studies to be reported here focused on spectral properties of the stimuli, particularly on the role played by the vocalic formant transitions. In Experiment I, we examined whether total elimination of vocalic formant transitions reduces the vowel context effect. In Experiment II, we dissociated effects on fricative perception due to vowel quality per se from effects due to the vocalic formant transitions, and we tested whether these two effects are differentially reduced by temporal separation of fricative noise and vocalic portion. In addition, each experiment dealt with a side issue: Experiment I, with the influence of a preceding vowel on fricative perception; Experiment II—a study using natural speech--with perceptual normalization induced by a difference in speaker.

EXPERIMENT I

In this study, we addressed the question of whether vocalic formant transitions must be present for a vowel to affect the perception of a preceding fricative. Since all our previous stimuli had contained formant transitions, the presumed effect of vowel context was confounded with whatever effect the transitions themselves might have had on fricative perception. Removal of formant transitions seemed one way of getting rid of this confounding and of assessing the contextual effect due to vowel quality per se.

A second question was that of the temporal order of fricative and vowel. This issue has been partially investigated in two earlier studies. Hasegawa and Daniloff (Note 2) synthesized a continuum of fricative noises ranging from [ʃ] to [s] and preceded these stimuli with either [i] or [u]. They found a reliable difference in the [ʃ]-[s] boundary between the two vocalic contexts. In contrast, Kunisaki and Fujisaki (Note 1), who embedded fricative noises between two vowels, found only a rather small (nonsignificant?) influence of the preceding vowel ([e,i,u,o,a]). In the present study, we compared the context effects of preceding and following (transitionless) vowels, separately and in concert.

Method

Subjects. The subjects included nine paid volunteers recruited from Yale University, a research assistant, and the two investigators. All had previously participated in Experiment I of Mann and Repp (1979a).

INFLUENCE OF VOCALIC CONTEXT ON PERCEPTION OF THE [ʃ]-[s] DISTINCTION: II. SPECTRAL FACTORS

Bruno H. Repp and Virginia A. Mann

Abstract. The position of the category boundary along a synthetic [ʃ]-[s] noise continuum depends on the following vocalic segment: When the vowel is [u], more "s" responses are given than when the vowel is [a]. In Experiment I, we showed that this context effect disappears when the vocalic portion is synthesized so as to contain no formant transitions. To dissociate the contribution of formant transitions from contextual effects due to vowel quality per se, Experiment II employed synthetic fricative noises followed by vocalic portions excerpted from naturally produced [ʃa], [sa], [ʃu], and [su]. The results showed strong and largely independent effects of formant transitions and vowel quality on fricative perception. Both effects were substantially reduced when a silent gap was inserted between the fricative noise and the vocalic portion, leading to perception of an intervening stop consonant; however, formant transitions continued to affect fricative perception even though they functioned now as cues to place of articulation of the stop. In this experiment, we provided additional evidence that the formant transitions following naturally produced [ʃ] and [s] are different from each other in perceptually significant ways, and we also found a strong speaker (male vs. female) normalization effect in fricative perception. Surprisingly, the speaker effect, too, was reduced by the presence of a silent gap.

INTRODUCTION

When synthetic fricative noises from a [ʃ]-[s] continuum are followed by [a] or [u] (with appropriate formant transitions), listeners perceive more instances of [s] in the context of [u] than in the context of [a] (Kunisaki & Fujisaki, Note 1). Presumably, this reflects a perceptual adjustment for the coarticulatory effect that rounded vowels have on preceding fricatives (through anticipatory lip rounding). In a recent investigation (Mann & Repp, 1979a), we replicated the basic perceptual effect and collected acoustic data from one speaker to corroborate the presence of an analogous coarticulatory effect in production. We also found that varying the duration of the

---

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories, and by NICHD Postdoctoral Fellowship HD05677 to Virginia Mann. We thank Alvin Liberman for his advice at all stages of this project, Doug Whalen for sharing his observations and insights with us, and Sarah Peck for assistance in data analysis.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

Stimuli. The stimulus materials were highly similar to those employed in Experiment I of Mann and Repp (1979a), and the reader is referred to that earlier paper for details. The primary stimuli were nine synthetic fricative noises that formed a [ʃ]-[s] continuum. Their duration was 150 msec. There were five conditions:

(1) Isolated fricative noises, presented in five randomized blocks of 21 stimuli with ISIs of 2.5 sec. (The number 21 resulted from a 1-2-3-3-3-3-3-2-1 frequency distribution of the nine stimuli on the continuum.)
(2) The same noises immediately followed by either [ta] or [tu], i.e., synthetic vocalic portions containing formant transitions roughly appropriate for a dental-alveolar place of articulation (and perceived as beginning with "d" in isolation). Because of the absence of a preceding silent interval, the vocalic formant transitions did not give rise to stop percepts; therefore, the vocalic context in this condition will be represented as [(t)a] and [(t)u]. Subjects heard [ʃa], [ʃu], [sa], or [su]. There were five blocks of 42 stimuli.
(3) The fricative noises followed by either [a] or [u], which were steady-state vowels obtained by straightening all formant transitions in [ta] and [tu]. Otherwise, this condition was identical to condition 2.
(4) The fricative noises preceded by either [a] or [u]. The initial vocalic portion was 150 msec in duration, with initial and final amplitude ramps but no formant transitions. Otherwise, this condition was identical to conditions 2 and 3.
(5) The fricative noises preceded and followed by either [a] or [u] without formant transitions. Thus, there were four combinations of vocalic context: [a-a], [a-u], [u-a], and [u-u]. This doubled the number of stimuli; they were presented in 5 blocks of 84.

Procedure. All subjects listened to the conditions in the same fixed order. The task was to identify the fricative consonant as "sh" or "s".

Results

The results are depicted in Figure 1 as the percentage of "sh" responses given to each stimulus along the fricative noise continuum. Figure 1a shows that condition 2 successfully replicated the basic context effect of the following vowel on perception of the fricative noise: There were fewer "sh" responses (hence, more "s" responses) in [-(t)u] context than in [-(t)a] context, $F(1,11) = 51.7$, $p < .0001$. As in our earlier study (Mann & Repp, 1979a: Exp. I), the effect was almost exclusively due to [-(t)u]. Perception of fricative noises in [(t)a] context was similar to their perception in isolation (condition 1, dotted function in Figure 1a). This asymmetry is in agreement with our interpretation of the context effect as reflecting perceptual compensation for effects of anticipatory lip rounding in production; such lip rounding, of course, is associated with [u] but not with [a].

Figure 1b shows what happened when the vocalic formant transitions were removed (condition 3). The context effect practically disappeared and was no longer significant. Curiously, however, the subjects gave fewer "sh" responses to noises followed by either vowel than to the noises in isolation.

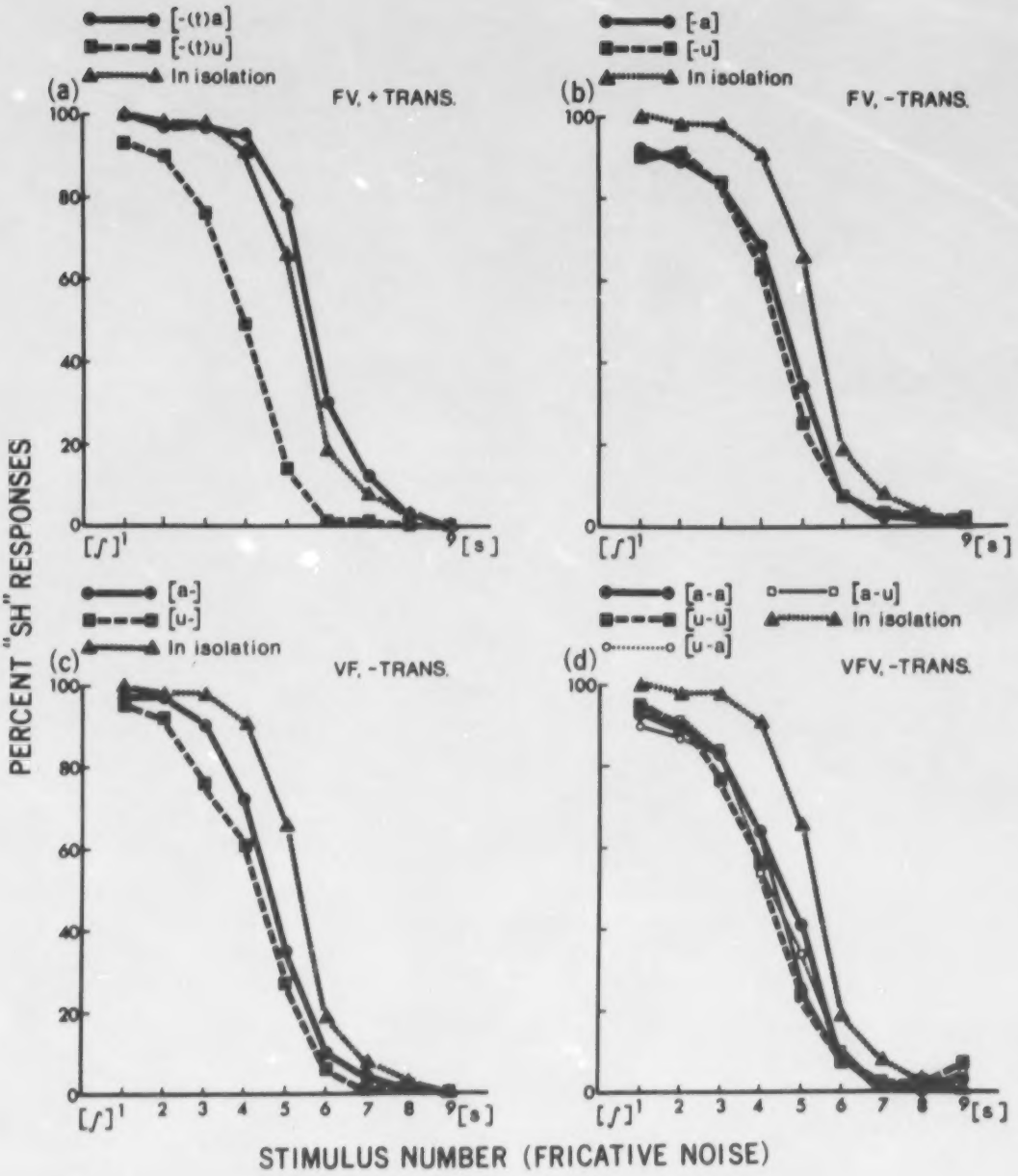


Figure 1. Effects of vocalic context on percentages of "sh" responses to stimuli from a fricative noise continuum in four conditions: fricative-vowel (FV) stimuli with formant transitions (+TRANS.); FV, -TRANS.; VF, -TRANS.; and VFV, -TRANS. The response function for isolated fricative noises (triangles, dotted) is replicated in all four panels.

Figure 1c shows that the context effect due to a preceding vowel (condition 4) was small, although it reached significance, $F(1,11) = 6.6$, $p < .05$: Subjects gave somewhat more "sh" responses to noises preceded by [a] than to noises preceded by [u]. However, the combined effects of preceding and following vowels (condition 5, Figure 1d) fell short of significance. In conditions 4 and 5, as in condition 3, there was a reduction in "sh" responses relative to the baseline rate for isolated noises.[1] The reason for this shift in response criteria is not known.

Discussion

Why did removal of the formant transitions from a following vowel eliminate its effect on the fricative? There are two possible answers: (1) The transitions held the fricative noise and the periodic portion together and in this way mediated the effect of the vowel on the fricative. (2) Alternatively, the transitions themselves, rather than the steady-state vowel portions, were the source of the context effect.

There is phenomenological evidence in support of the first explanation. Elimination of the formant transitions resulted in a less coherent stimulus percept. In the authors' perception, the fricative noise seemed to be segregated from the vocalic portion and to come from a different source. Thus, the pattern sounded less speechlike, which may also be the reason why fricatives followed by steady-state vowels show no right-ear advantage in dichotic presentation (Darwin, 1971). According to this interpretation, the aperiodic and periodic portions of the stimulus must be perceptually coherent for a context effect to arise. Appropriate formant transitions seem to establish auditory (and perceived articulatory) continuity between fricative noise and vocalic portion in the same way that they make successive vowels hang together (Dorman, Cutting, & Raphael, 1975). The disappearance of the context effect in transitionless fricative-vowel stimuli is in agreement with the similar absence of perceptual interactions between stimuli with different apparent sources due to different fundamental frequencies (Ades, 1977; Darwin & Bethell-Fox, 1977; Dorman, Raphael, & Liberman, 1979).

In contrast to following transitionless vowels, preceding transitionless vowels seemed to exert a small effect on fricative perception. It is interesting to note, in that connection, that the perceptual coherence of vowel-fricative stimuli was much less disrupted by the absence of formant transitions than the perceptual coherence of fricative-vowel stimuli, perhaps because transitions preceding fricatives are less extensive in natural speech or less salient in perception than transitions following fricatives. (Note the similar difference in perceptual salience between transitions preceding and following closure of intervocalic stops--Repp, 1978.) The presence of a small context effect in vowel-fricative stimuli supports the hypothesis that perceptual coherence promotes context effects.[2]

Persuasive as these observations may be, we need to consider the possibility that the supposed vowel context effect in fricative-vowel stimuli was actually due to the formant transitions themselves acting as cues to place of articulation of the fricative. Perhaps, the transitions in [(t)u] were relatively more appropriate for [s] than those in [(t)a]. Fortunately, there

is indirect evidence (presented in connection with Exp. II of Mann & Repp, 1979a) that this was not the case. Nevertheless, it seemed important to determine directly the relative contributions of formant transitions and vowel quality to the context effect. This was the purpose of Experiment II.

EXPERIMENT II

Our second experiment had the primary purpose of dissociating the contributions of formant transitions (as place-of-articulation cues) and vowel quality to the vowel context effect. To that end, it was necessary to vary these two factors independently. When using synthetic speech, one can never be sure that transitions in different vocalic portions (such as [ta] and [tu]) are equally appropriate (or equally neutral) for one or the other fricative place of articulation. Therefore, we decided to take the vocalic portions from natural utterances of fricative-vowel syllables and to combine them with our synthetic fricative noises. In this way, we could be assured that the formant transitions were indeed appropriate for either [ʃ] or [s], depending on the original utterance.

For a long time, the contribution of vocalic formant transitions to the [ʃ]-[s] distinction was considered negligible (Harris, 1958; LaRiviere, Winitz, & Herriman, 1975). However, recent studies by Whalen (1979) suggest that the transitions do contribute to the distinction once the noise cue is weakened or neutralized. Whalen used stimuli along a synthetic fricative noise continuum similar to ours and followed them with synthetic or natural vocalic portions containing transitions either more appropriate for [ʃ] or more appropriate for [s]. He demonstrated clear effects of the transitions; these effects were especially large in the case of natural vocalic portions, presumably because these natural signals either contained additional cues to place of articulation or because the transitions were given more perceptual weight because of their naturalness. Whalen also found an effect of vowel quality that was largely independent of the transition effect. Thus, he anticipated the effects we hoped to find in the present study. However, his vowels varied from [i] to [u]; they did not include [a]. To enable us to make a more direct comparison with our earlier studies, we replicated and extended Whalen's experiments using the vowels [a] and [u].

Experiment II extended Whalen's studies by including a condition in which the vocalic portion and the fricative noise were separated by an 87-msec silent gap, which led to perception of fricative-stop-vowel syllables. Thus, we examined whether the transition and vowel quality effects show different amounts of reduction as a consequence of temporal separation. We expected that the transitions would contribute to fricative perception only as long as they were interpreted as cues to fricative place of articulation; as soon as a stop consonant was heard, they would be understood as cues to place of articulation of the stop and lose their effect on the fricative. Thus, we predicted that the transition effect--which really is not a "context effect" at all, but a perceptual effect due to integration of cues for a given phonetic segment (cf. Repp, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978)--would be truly eliminated when an intervening phonetic segment (viz., a stop consonant) is heard; whereas the vowel quality effect, which depends only

on temporal separation, might still be present at an 87-msec gap, although in reduced form (cf. Mann & Repp, 1979a: Exp. II).

A significant differential effect of formant transitions on fricative perception when no silent gap intervenes would be sufficient evidence that the vocalic transitions following naturally produced [ʃ] and [s] noises are indeed different from each other, as one might expect from the different places of articulation of the two fricatives. We did not attempt to assess this difference directly, since formant transitions are notoriously difficult to measure in spectrograms. Instead, we adduced two additional kinds of perceptual evidence. One was the place of articulation assigned to the stop consonants in the gap condition; we expected that natural [ʃ]-transitions, which reflect a relatively more posterior place of articulation, would lead to more "k" (and fewer "t") responses than natural [s]-transitions. In addition, we presented the isolated vocalic portions in a separate condition and asked the subjects to identify the initial stop consonant, if one was heard. We expected to find a similar pattern of place-of-articulation identifications here as in the gap condition.

One additional feature of Experiment II must be mentioned before we proceed to describe the method of this multifactorial study in more detail. In order to assure that our results would not be specific to the particular tokens selected from natural utterances, we not only used multiple tokens but also two speakers, one male and one female. Consequently, the vocalic portions that followed our synthetic fricatives reflected different vocal tract sizes and source characteristics. We wondered whether these differences (hereafter referred to collectively as the speaker difference) would influence fricative perception in the no-gap condition, and whether this effect would persist in the gap condition. That there are detectable acoustic differences between the fricative noises produced by males and females has been shown by Schwartz (1968). The spectra of female [ʃ] and [s] noises are shifted upwards on the frequency scale, relative to those produced by males, presumably because of differences in vocal tract size. We might expect that speaker-specific information conveyed by a vocalic portion would lead listeners to change their criteria in deciding on the preceding fricative, such that the [ʃ]-[s] boundary on a synthetic noise continuum is shifted towards higher frequencies in the context of a female voice. Indeed, precisely such a perceptual normalization effect has been reported by May (1976) who followed synthetic fricative noises with synthetic vocalic portions whose formants were scaled upward or downward to simulate changes in vocal tract size. Our Experiment II was intended to confirm May's finding with natural-speech vocalic portions. We were particularly curious to see whether the speaker effect, if obtained, would decline as a gap was inserted. If it did, this would have some interesting implications for a theory of perceptual normalization.

In summary, then, this study examined the effects of orthogonal variations in three parameters--formant transitions, vowel quality, and speaker characteristics--on fricative perception, as well as changes in each of these effects consequent upon introduction of a gap (and a stop consonant percept) between fricative and vowel.

Method

Subjects. The subjects in the main experiment were six paid volunteers, a research assistant, and the two authors. Three additional subjects had to be eliminated because they had difficulties in the gap condition.[3] Five new paid volunteers, the research assistant, and the two authors listened to the isolated vocalic portions.

Stimuli. Two adults, one male and one female, both native speakers of American English, spoke the utterances [ʃa], [ʃu], [sa], [su] ten times in a random sequence that included several other utterance types. All utterances were recorded on magnetic tape in an anechoic chamber and subsequently digitized at 10kHz using the Haskins Laboratories Pulse Code Modulation (PCM) system. Aided by our ears and by waveform displays, we selected three good tokens of each utterance for use in the experiment. There were 24 stimuli altogether (2 speakers x 2 fricatives x 2 vowels x 3 tokens). Using the PCM computer programs, the fricative noises (defined as the signal portion preceding the first detectable pitch pulse in an oscillogram) were removed from all stimuli, and the digitized synthetic fricative noises from our nine-member continuum were substituted instead.[4] This led to a total of 216 (9 x 24) stimuli--recorded in two completely random sequences with ISIs of 3 sec and a 6-sec ISI after each group of 24. A second set of stimuli was constructed by inserting an 87-msec period of silence between the synthetic fricative noises and the natural vocalic portions. These stimuli were recorded in identical sequences.

A few changes were made in the synthetic fricative noises, designed to improve their naturalness. In part, these changes were modelled after the natural fricative noises produced by the male speaker in our experiment. The spectral structure of the noises remained unchanged (see Table I in Mann & Repp, 1979a). However, they were extended to 200 msec duration and received a new, presumably more realistic, amplitude contour. The contour was triangular: Amplitude increased over the first 150 msec and decreased over the final 50 msec. To insure that all noises had approximately the same overall amplitude while retaining essentially the same amplitude contour, we specified equal amplitudes for all noises at the synthesis stage, digitized the resulting output that varied by about 12 dB ([ʃ] < [s]) between the extremes of the noise continuum (a consequence of hardware constraints), and then adjusted the amplitudes of the digitized noises in eight 1-dB steps, which reduced the effective amplitude difference between the continuum end-point stimuli to about 4 dB.[5]

The natural vocalic portions varied considerably in duration (300-500 msec) but were relatively homogeneous in amplitude. The [-a] portions were approximately equally intense for the two speakers. The [-u] portions of the female speaker were lower by about 5 dB; those of the male speaker, by only about 1 dB. An additional stimulus tape was recorded containing five randomized sequences of these 24 isolated vocalic portions.

The amplitudes of the synthetic fricative noises were 13-21 dB below the natural vocalic portions, depending on the individual token. It was interesting to note that, for both of our speakers, the amplitudes of the natural [ʃ]

noises in the original utterances had been 8-11 dB higher than those of the [s] noises; they had been an average of 8 and 17 dB, respectively, below the natural [-a] portions. This difference directly contradicted the amplitude gradient built into the OVE IIIc synthesizer, which gave the [s] noise a higher intensity. By (approximately) amplitude-equalizing our synthetic noises, as described above, we reached a reasonable compromise.

Procedure. Some subjects listened twice to the no-gap tape before listening twice to the gap tape in a separate session. Others were presented with the no-gap tape followed by the gap tape in each of two sessions. A total of four responses to each individual stimulus was obtained from each subject, or twelve responses when ignoring token differences. The task in the no-gap condition was to identify the fricatives as "sh" or "s". In the gap condition, the following stop consonant (if heard) had to be identified as well; the relevant response choices were "p", "t", "k", or "-" (no stop), or whatever other stop-like sounds might be heard.

The tape with the isolated vocalic portions was presented to a partially different group of subjects whose task was to identify the initial stop consonant (if heard); the suggested response choices were "b", "th" (for the initial sound in that), "d", "g", "-" (no stop), but the subjects were encouraged to write down any other consonantal sounds they heard.

Results

Fricative Identification. The results of the main experiment are shown in Figures 2 and 3. Figure 2 displays percentages of "sh" responses as a function of stimulus position along the fricative noise continuum. The panels on the left display the no-gap condition, those on the right the gap condition. Top panels are for the male speaker, bottom panels for the female speaker. The four different functions in each panel correspond to the four original utterances, [ʃa], [ʃu], [sa], and [su], from which the vocalic portions derived. The data have been averaged over the nine subjects and over the three different tokens of each vocalic portion. Figure 3 summarizes these data in terms of the overall percentage of "sh" responses (averaged over the nine stimuli on the fricative noise continuum) as a function of original utterance. This figure displays the token variation (data point triplets) and makes the speaker effect (male vs. female) easier to see.

In the no-gap condition, the [ʃ]-[s] distinction was strongly affected by all three factors investigated: vowel quality, $F(1,8) = 35.7$, $p < .001$; formant transitions, $F(1,8) = 52.6$, $p < .001$; and speaker, $F(1,8) = 52.7$, $p < .001$. Listeners gave substantially more "sh" responses to fricative noises followed by [-a], or [ʃ] transitions, or a female voice, than to noises followed by [-u], or [s] transitions, or a male voice. Expressed as overall differences, the transition effect was the largest (42 percent), followed by the vowel quality (25 percent) and speaker (22 percent) effects. All three effects were in the predicted direction and so strong that seven out of eight response functions (Fig. 2, left panels) did not reach asymptote at both ends. For example, when the vocalic portion derived from a female [ʃa], even the most [s]-like noise received 65 percent "sh" responses; and when the vocalic portion derived from a male [su], even the most [ʃ]-like noise received only

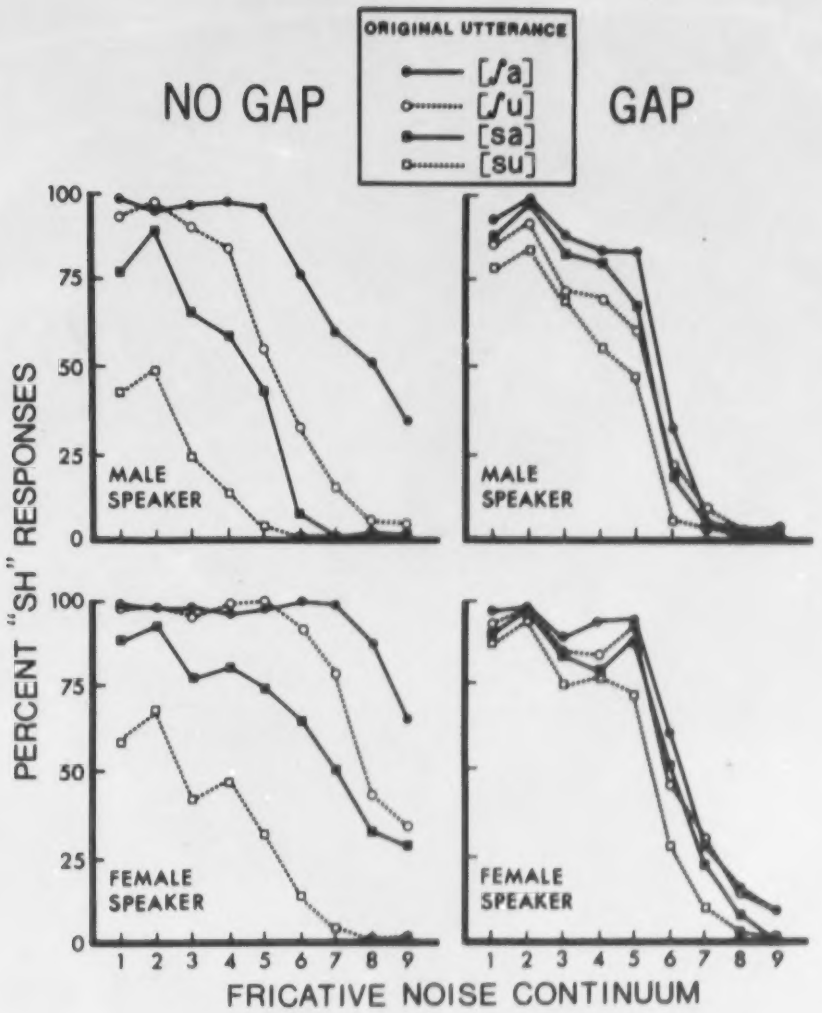


Figure 2. Effects of natural formant transitions and vowel quality on "sh" responses to stimuli from a synthetic fricative noise continuum, for two speakers (male, female) in two conditions (no-gap, gap).

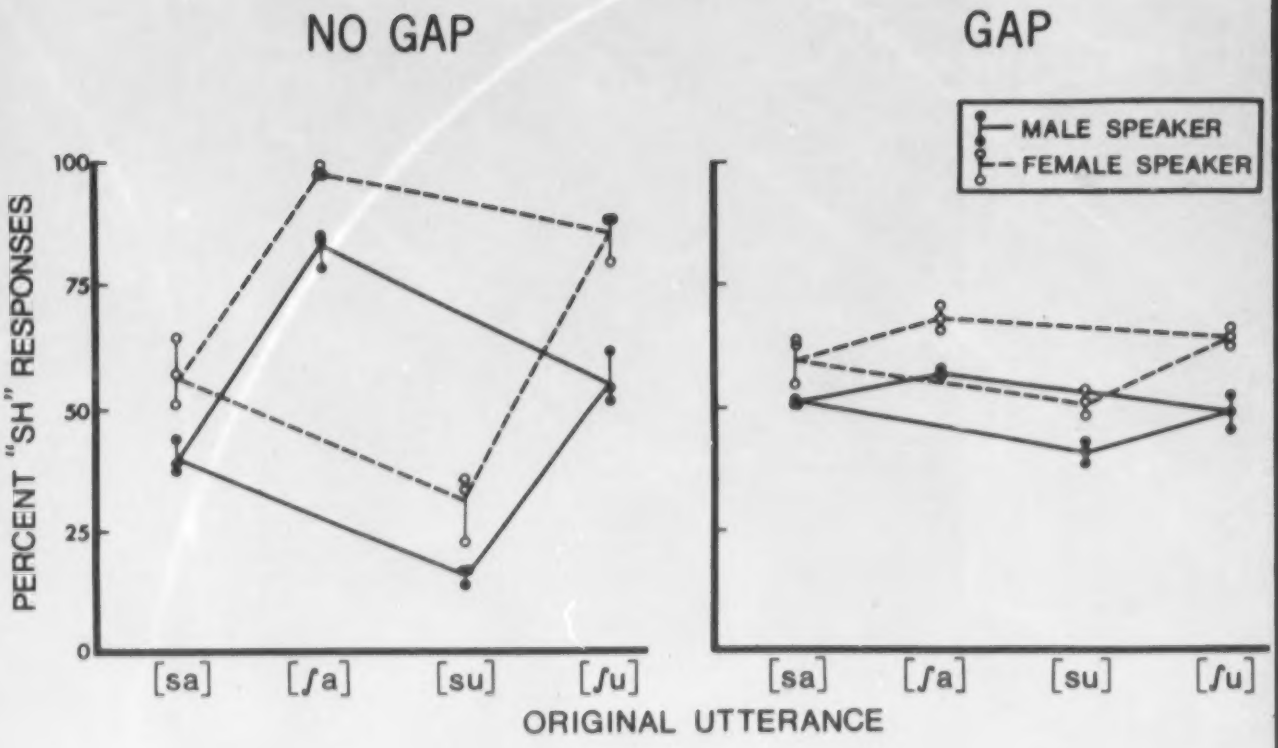


Figure 3. Summary of Figure 2 in terms of overall percentages of "sh" responses, with token variation displayed. Ascending lines--transition effect; descending lines--vowel quality effect; vertical displacement--speaker effect.

50 percent "sh" responses. The transition effect was larger with [-u] than with [-a], $F(1,8) = 41.7$, $p < .001$; this interaction was more pronounced with the female voice than with the male, $F(1,8) = 6.9$, $p < .05$.

Consider now the results of the gap condition, shown in the right-hand panels of Figures 2 and 3. As can be seen, all experimental effects were substantially reduced here. All response functions were close to asymptote at either end and had similar shapes, with a major drop in "sh" responses between stimuli 5 and 6 on the noise continuum, which is in accord with the identification of these noises in isolation (cf. Fig. 1a). The decline in magnitude consequent upon introduction of an 87-msec gap was significant for all three effects: vowel quality, $F(1,8) = 11.8$, $p < .01$; formant transitions, $F(1,8) = 41.6$, $p < .001$; and speaker, $F(1,8) = 10.8$, $p < .02$. The reduction of the transition effect was not only most reliable statistically but also the largest numerically (25 percent vs. 13 percent for vowel quality and 10 percent for speaker). Nevertheless, all three effects were still present in the gap condition: vowel quality, $F(1,8) = 12.9$, $p < .01$; formant transitions, $F(1,8) = 14.9$, $p < .01$; and speaker, $F(1,8) = 18.5$, $p < .01$. Expressed as overall differences, the transition effect was still the largest (17 percent), followed by speaker (12 percent) and vowel quality (8 percent). There was no longer any interaction between the transition and vowel quality effects.

It is evident from Figure 3 that token variation was small relative to the effects under investigation, even though some token differences appeared to be systematic and reliable. An analysis of variance with token variance as the error term yielded essentially the same results as the earlier analysis (which used treatment-by-subject interactions as error terms). A min F' analysis (Clark, 1973), combining subject and token variability, again yielded similar results; in particular, the three main effects remained significant at the $p < .001$ level in the no-gap condition and at the $p < .01$ level in the gap condition.

Stop Consonant Identification. Let us now examine how the vocalic formant transitions were perceived when they cued place of articulation of a stop consonant. Consider first the results of the condition in which the isolated vocalic portions were presented for identification. Responses were obtained in each of the suggested categories: "b", "th" (as in that), "d", "g", and "-" (no initial consonant heard). Three additional categories, "h", "y", and /ɣ/ were contributed by two subjects, one of whom had had phonetic training. The latter responses constituted only 5 percent of the total, and since they designated relatively posterior places of articulation and exhibited a pattern similar to that of the more frequent "g" responses, they were combined with "g" responses for purposes of analysis. The percentages of responses in this enlarged "g" category are displayed in the left panel of Figure 4, in a manner analogous to Figure 3. The percentages in all response categories, averaged over tokens, are listed in the left half of Table 1.

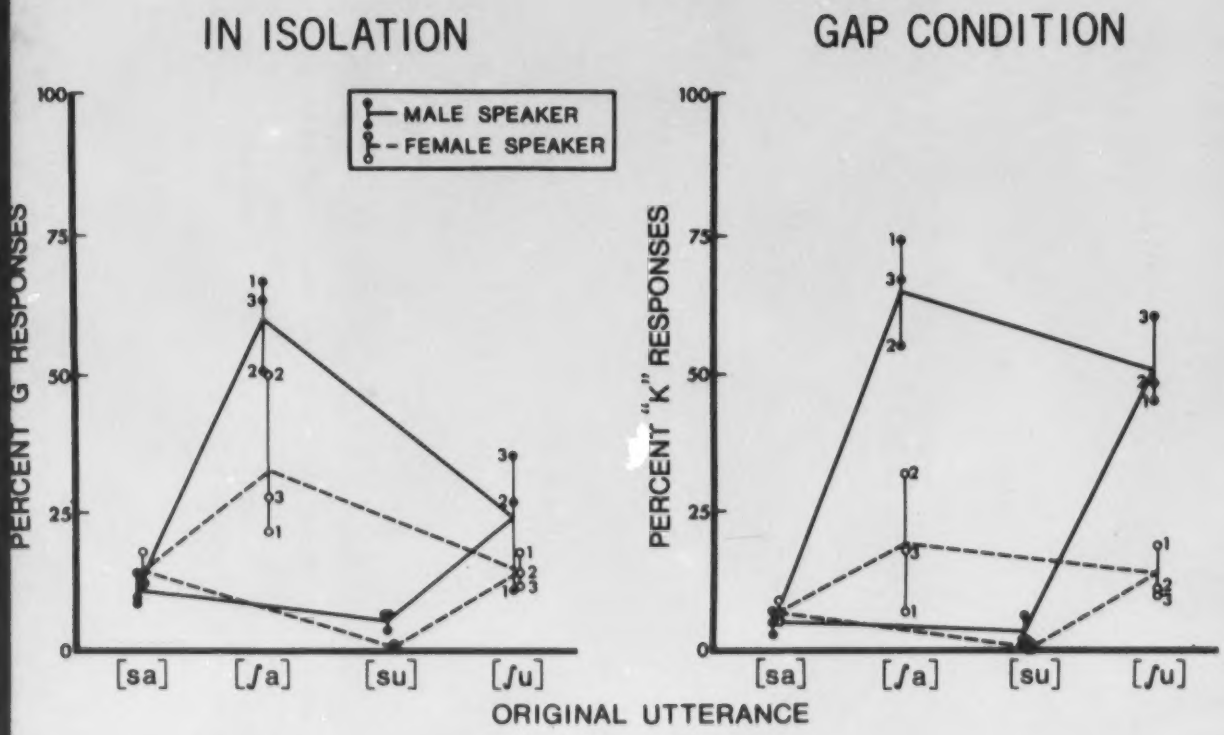


Figure 4. Overall percentages of velar stop responses to vocalic portions in isolation and when preceded by a fricative noise plus gap. Small digits indicate individual tokens (arbitrarily numbered).

Table 1

Stop consonant identification in isolated vocalic portions and when preceded by a fricative noise plus gap.

Original utterance	Responses (in percent) In isolation b	th	d	g	-	Gap condition p	t	k	-
Male speaker									
[sa]	17.1	47.9	24.2	9.2	1.7	23.8	75.3	5.1	4.2
[ʃa]	0.4	10.0	24.2	60.1	5.4	2.4	27.9	65.3	4.4
[su]	4.6	34.2	35.0	5.5	20.9	2.3	74.3	3.4	20.0
[ʃu]	0.4	2.1	45.9	31.3	20.5	8.2	27.6	51.1	13.1
Female speaker									
[sa]	13.4	53.4	17.5	14.2	1.7	9.0	79.2	6.6	5.2
[ʃa]	0.4	19.6	43.3	32.9	3.8	0.9	73.8	19.3	6.0
[su]	5.9	34.6	56.3	0.4	2.5	0.1	80.9	1.1	17.9
[ʃu]	2.5	13.4	55.9	14.2	14.2	2.2	54.5	13.6	29.7

Figure 4 shows that the expected difference between [ʃ] and [s] transitions was reflected in perception. Vocalic portions that had originally been preceded by [ʃ] were more often perceived as beginning with a velar stop than vocalic portions that had originally been preceded by [s]. Thus, the formant transitions associated with the more posteriorly articulated fricative, [ʃ], favored stop percepts with a posterior place of articulation.6 Conversely, Table 1 shows that [s] transitions received more "b" and "th" responses than [ʃ] transitions, which again confirms our expectations, since [s] has a relatively anterior place of articulation.

Turning now to the identification of the same vocalic portions in the gap condition (i.e., when preceded by a fricative noise plus 87 msec of silence), we find much the same pattern with increased regularity. In the right panel of Figure 4, we see that more "k" responses were given to [ʃ] transitions than to [s] transitions, and that this difference was particularly pronounced for the male voice. (There were also somewhat more "k" responses in [-a] context than in [-u] context.) Token variability was large, and the rank orders of

tokens were identical with those obtained in isolation, as can be seen by comparing the two panels of Figure 4. In the right half of Table 1, we see that "p" and "t" responses (the "th" category, of course, was not appropriate for consonants following a fricative) were more frequent to [s] transitions than to [ʃ] transitions, especially for the male speaker. Unexpectedly, there was a fair percentage of "no stop" responses in the gap condition, and these responses occurred primarily in [-u] context, as they did also for isolated vocalic portions.

Discussion

The results of the no-gap condition demonstrate that the [ʃ]-[s] distinction is affected both by the quality of the following vowel and by the vocalic formant transitions. Our findings are in excellent agreement with those of Whalen (1979), including the large size of the effects that was presumably due to increased perceptual weights of natural acoustic cues (contained in the vocalic portion) in relation to synthetic cues (fricative noise). Whalen showed that both effects are reduced in size in all-synthetic stimuli. His, and our, successful experimental dissociation of the vowel quality effect from the transition effect reinforces our earlier conclusion that the "vowel context effect" obtained in Experiment I (and by Kunisaki & Fujisaki, Note 1; Mann & Repp, 1979a) was indeed primarily due to vowel quality, even though the formant transitions had not been varied orthogonally.

Despite their paradigmatic similarity, the vowel quality and formant transition effects are very different phenomena from a theoretical perspective. The formant transitions are a consequence of the articulatory movements involved in producing the fricative consonant. Thus, they constitute a perceptual cue to fricative place of articulation; this cue is integrated with others (such as the fricative noise) into a unitary phonetic percept. Vowel quality, on the other hand, is neither a consequence of fricative production, nor a direct cue to fricative perception. Rather, it is an independent factor that affects the production of the fricative, and this effect is somehow compensated for in perception. Thus, only the vowel quality effect is a true context effect; the transition effect is a manifestation of perceptual cue integration (cf. Repp, 1978; Repp et al., 1978).

It was this theoretical distinction that led us to predict that the vowel quality and transition effects would be differentially affected by insertion of a silent gap between fricative noise and vocalic portion. We hypothesized that when the transitions are interpreted as cues to place of articulation of a stop consonant they would lose their effect on fricative perception. The vowel quality effect, on the other hand, was expected to persist in reduced form, since this was the result obtained in earlier studies (Mann & Repp, 1979a). These predictions were only partially confirmed. The vowel quality effect was indeed reduced, and the transition effect even more so. However, in addition to a diminished vowel quality effect, a significant transition effect persisted in the gap condition. This effect cannot be accounted for by the fact that no stop consonants were heard on some trials, despite the gap. For example, the subject with the largest transition effects in the gap condition always heard stops.

Persistence of the transition effect could be explained post hoc in at least two ways. One is that perceptual integration of the fricative noise and transitional cues was not blocked despite the perception of an intervening stop consonant. This may have occurred because some cues that normally block integration were absent, most notably the plosive burst following the stop closure. After all, the stimuli in the gap condition derived from fricative-vowel utterances, not from fricative-stop-vowel utterances. Therefore, the nature of the acoustic cues may have promoted integration, and the perception of an intervening stop may have been a mere epiphenomenon. Another possibility is that there is a perceptual dependency between a fricative and a following stop consonant, such that listeners are more likely to hear [s] when [t] follows and [ʃ] when [k] follows. However, this interpretation seems not only less plausible on both perceptual and articulatory grounds; we also have evidence to the contrary: Mann and Repp (1979b) found that [t] and [k] affect the perception of preceding [ʃ] or [s] in precisely the opposite way. Therefore, we opt for the first interpretation--that the acoustic cues, because of their origin in fricative-vowel utterances, promoted cue integration despite perception of an intervening phonetic segment. In other words, the formant transitions contributed to the perceived place of articulation of two segments--the fricative and the stop consonant.

By demonstrating a strong differential effect of [ʃ] vs. [s] transitions on fricative perception, the present study provided strong evidence that these transitions were, in fact, different from each other. To this indirect perceptual evidence must be added the direct perceptual evidence from stop consonant identification in stop-vowel and fricative-stop-vowel stimuli: [ʃ] transitions were more often associated with a posterior (velar) place of stop articulation than [s] transitions, and [s] transitions were more often associated with an anterior (labial or dental) place of stop articulation than [ʃ] transitions. Although these types of perceptual evidence cannot replace a traditional spectrographic analysis, they are more meaningful for two reasons: A large difference visible in spectrograms may have little perceptual importance, whereas a difference that is difficult to discern spectrographically may be perceptually significant. Since degree of perceptual salience is the most interesting aspect of acoustic cues, direct and indirect perceptual evidence of the sort provided here answers the question about differences between [ʃ] and [s] transitions in a theoretically more satisfying way than spectrographic measurements would. The indirect method, especially, suggests itself as a powerful procedure for the perceptual assessment of coarticulatory effects in natural speech. We are in the process of exploiting it in several other contexts.

The perceptual assessment of the transitional cues revealed additional, unexpected differences that raise interesting questions. For example, it was found that the [ʃ] transitions of the male speaker led to many more "k" (or "g") responses than those of the female speaker (see Figure 4). This suggests a difference between the two speakers in the way in which [ʃ] was articulated. Moreover, within each speaker there seemed to be systematic variability between individual tokens of [ʃ] and [s] transitions, again indicating systematic articulatory variability. Thus, the method of perceptual assessment of natural-speech cues may reveal hitherto unsuspected variation in (fricative) articulation that warrants further investigation.

A final comment is necessary concerning the effect of speaker characteristics on fricative perception. We found that, in the no-gap condition, listeners gave substantially more "sh" responses to noises in context of the female voice. This result confirms May (Note 4) and suggests that listeners compensate, or "normalize", for changes in fricative noise spectrum induced by differences in vocal tract size. The extent of the perceptual compensation seemed to be much larger than actual differences in fricative spectrum between male and female speakers (Schwartz, 1968). A similar observation was made for the vowel context effect (Mann & Repp, 1979a). Thus, listeners seem to overcompensate (or "hypernormalize") in perception. The most interesting finding, though, was that the speaker effect decreased substantially with introduction of an 87-msec gap. This finding has important theoretical implications: It indicates a divergence of perception and production. Effects of vocal tract size on fricative noise spectrum should be independent of the context in which the fricative occurs; however, perceptual compensation for such effects proves to be context-dependent. It may be argued that the speaker effect on fricative perception was reduced in the gap condition because the fricative noise was perceptually dissociated from the vocalic portion, as if these two components did not belong to the same utterance. Yet, this does not provide an explanation, it merely describes the possible phenomenological consequence of introducing a gap. The important implication of the result is that perceptual normalization effects are sensitive to local temporal properties of the speech signal. In this way they seem to be rather similar to perceptual effects of speaking rate (Summerfield, Note 3; Miller, in press). In each case the perceptual effects seem to operate only over a limited temporal region, suggesting the involvement of a rapidly decaying auditory memory or a sliding perceptual integrator with a time window of a few hundred milliseconds.

To summarize: We have demonstrated that an effect of vowel quality on perception of a preceding fricative exists independently of effects due to the vocalic formant transitions (Exp. II), although formant transitions appear to be necessary to preserve the perceptual coherence of the utterance and thus enable vowel quality to have its perceptual effect (Exp. I). We have provided three-fold evidence that the vocalic formant transitions following naturally produced [ʃ] and [s] are different from each other in perceptually significant ways. Finally, we have shown that perception of fricative noises conforms to characteristics of the vocal tracts that listeners believe have produced those noises--and that this normalization effect, like the effects of vowel quality and formant transitions, is sensitive to local temporal properties of the stimulus.

REFERENCE NOTES

1. Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. Annual Bulletin of the Research Institute for Logopedics and Phoniatrics (University of Tokyo), 1977, 11, 85-91.
2. Hasegawa, A., & Daniloff, R. G. Effects of vowel context upon labelling the /s/-/š/ continuum. Paper presented at the 91st Meeting of the Acoustical Society of America, Washington, D. C., April 1976

(unpublished).

3. Summerfield, Q. On articulatory rate and perceptual constancy in phonetic perception. Unpublished manuscript, 1977.

REFERENCES

Ades, A. E. Source assignment and feature extraction in speech. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 673-685.

Bell-Berti, F., & Harris, K. S. Anticipatory coarticulation: Some implications from a study of lip rounding. Journal of the Acoustical Society of America, 1979, 65, 1268-1270.

Bondarko, L. V. The syllable structure of speech and distinctive features of phonemes. Phonetica, 1969, 20, 1-40.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Darwin, C. J. Ear differences in the recall of fricatives and vowels. Quarterly Journal of Experimental Psychology, 1971, 23, 46-62.

Darwin, C. J., & Bethell-Fox, C. E. Pitch continuity and speech source attribution. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 665-672.

Dorman, M. F., Cutting, J. E., & Raphael, L. J. Perception of temporal order in vowel sequences with and without formant transitions. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 121-129.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 1979, 65, 1518-1532.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L J. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. Perception and Psychophysics, 1977, 22, 109-122.

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1958, 1, 1-7.

LaRiviere, C., Winitz, H., & Herriman, E. The distribution of perceptual cues in English prevocalic fricatives. Journal of Speech and Hearing Research, 1975, 18, 613-622.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction: I. Temporal factors. Haskins Laboratories Status Report on Speech Research, 1979, SR-59/60, 49-64. (a)

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. Haskins Laboratories Status Report on Speech Research, 1979, SR-59/60, 65-82. (b)

May, J. Vocal tract normalization for /s/ and /š/. Haskins Laboratories Status Report on Speech Research, 1979, SR-48, 67-74.

Miller, J. L. The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In P. D. Eimas & J. L. Miller (Eds.), Perspectives on the study of speech. Hillsdale, NJ: LEA, in press.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. Perception and Psychophysics, 1978, 24,

471-485.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 621-637.

Schwartz, M. F. Identification of speaker sex from isolated voiceless fricatives. Journal of the Acoustical Society of America, 1968, 43, 1178-1179.

Whalen, D. H. Effects of vocalic formant transitions and vowel quality on the English /s/-/š/ boundary. Haskins Laboratories Status Report on Speech Research, 1979, SR-59/60, 35-48.

FOOTNOTES

[1]A three-way analysis of variance of conditions 1 (with Context as a dummy variable), 3, 4, and 5 ([a-a] and [u-u] contexts only), with the factors Preceding Vowel (V_1), Following Vowel (V_2), and Context ([a] vs. [u]) revealed significant effects of V_1, V_2, and $V_1 \times V_2$, $F(1,11) = 10.3$, 8.9, and 8.2, respectively, all $p < .02$. All three effects were due to the higher number of "sh" responses to isolated noises.

[2]The small size of the context effect in our vowel-fricative stimuli was probably due to the absolute position of the vocalic portion--preceding rather than following the fricative noise--and not to the absence of formant transitions. Kunisaki and Fujisaki (Note 1) apparently specified transitions in their initial vowel portions (cf. their Figure 7), although no detailed description was given. Nevertheless, their context effect was similar in magnitude to ours. Moreover, Bondarko (1969) mentions in passing that preceding vowels have a smaller coarticulatory effect on fricatives than following vowels. More systematic spectrographic and articulatory data are needed on this issue, but Bondarko's observation is entirely in accord with the general hypothesis (Mann & Repp, 1979a, 1979b) that perceptual context effects reflect the listeners' implicit knowledge of articulatory dependencies.

[3]Of these three subjects, one heard too many instances of [ʃ] in the gap condition, as well as unusual following consonants, such as [n] and [l]. A second subject responded erratically to the fricative noises and heard many instances of [n]. The third subject heard no stop consonants at all (as in the no-gap condition) and showed a nearly random pattern of fricative responses. All three subjects, however, gave a regular pattern of results in the no-gap condition, similar to that exhibited by the other subjects. Their no-gap results were excluded to make possible a pure within-subject comparison of the no-gap and gap conditions.

[4]Due to a digitizing artifact which escaped our notice at the stage of stimulus construction, an unintended 12-msec interval of silence slipped in between fricative noise and vocalic portion in these stimuli. However, since this interval was not sufficient to lead to the perception of stop consonants, its presence was considered inconsequential. For expository reasons, we will continue to refer to this condition as the no-gap condition.

[5]This procedure was superior to that employed in Experiment II of Mann and Repp (1979a), where amplitude adjustments were made at the synthesis stage. One consequence of the earlier procedure was that the more [s]-like noises had progressively flatter amplitude contours and more abrupt onsets. While this could hardly have affected the [ʃ]-[s] distinction, it was an aesthetic flaw that could be remedied in the present experiment, where digitized stimuli were employed.

[6]Several other differences are evident in Figure 4: More "g" responses were given in [-a] context than in [-u] context; more "g" responses were given to the male voice than to the female voice; for the male speaker, there was a larger transition effect in [-a] context than in [-u] context; and there were considerable token differences in the perception of [ʃ] transitions. The first effect (vowel context) may derive from the fact that velar stops are difficult to hear in [-u] context when the stimuli do not contain any burst (Dorman, Studdert-Kennedy, & Raphael, 1977). The last effect, the apparently greater variability of [ʃ] transitions, is an artifact due to the low probability of "g" responses to [s] transitions; [s] transitions showed similarly large token variability when other response categories were considered.

EXPLORING A VIBRATORY SYSTEMS ANALYSIS OF HUMAN MOVEMENT PRODUCTION*

J. A. Scott Kelso+ and Kenneth G. Holt+

Perhaps the most desirable attributes of the human motor system are that it be capable of locating the limbs accurately in space using a variety of movement trajectories and that localization be accomplished relatively independent of changes in the initial conditions of the limbs. Although it is well-documented that these features are characteristic of the behavioral repertoire of both animals and humans, less clear is the nature of the underlying control mechanism(s). Neither of the currently popular closed-loop, feedback (Adams, 1977) or open-loop, programming accounts (Schmidt, 1976) seem completely adequate. For example, although a closed-loop model could accommodate the fact that achievement of final position is possible in spite of (a) changes in limb position prior to movement (Stelmach, Kelso, & Wallace, 1975) or (b) the introduction of abrupt changes in load during movement execution (Houk, 1978; Polit & Bizzi, 1978), it is at a loss when the same findings can be demonstrated under deafferentation conditions (Bizzi & Polit, in press; Kelso, 1977; Nashner & Grimm, 1978; Taub, Goldberg, & Taub, 1975). Similarly, central motor programs that do not require ongoing feedback monitoring may handle deafferentation findings, but go awry when confronted with unforeseen changes in the movement context. Indeed, even a hybrid model that incorporates internal, central feedback loops (Evarts, 1971; Kelso & Stelmach, 1976) has great difficulty with the finding that normal accuracy may result when monkeys are deafferented and consequently subjected to unpredictable movement perturbations (Polit & Bizzi, 1978).

An alternative approach to problems of localization--stemming from Bernstein's original work (1947)1, proposes that where muscles at a joint are constrained to act as a unit, the linkage is describable as a class of vibratory system with the physical and behavioral characteristics of a mass-spring (Asatryan & Fel'dman, 1965; Fel'dman, 1966; Polit & Bizzi, 1978). There are several properties of a mass-spring that are advantageous in explaining the style of control observed in localization experiments. Perhaps its major characteristic for our purposes is that it is intrinsically self-equilibrating; once set in motion the spring will always come to rest at the same resting length (equilibrium point). Neither an increase in initial deflection of the spring from its resting length nor temporary perturbations will prevent the achievement of the equilibrium point, a property known as

---

*To be published in Journal of Neurophysiology.

+Also at University of Connecticut, Storrs.

Acknowledgment: We thank James Pruitt for his assistance in data collection and two anonymous reviewers for their helpful remarks. This research was supported by NIH grants NS 13617 and AM 25814.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

equifinality (von Bertalanffy, 1973). Support for this account comes from experiments in which subjects were required to hold a steady angle at the elbow joint against a resistance and not to make adjustments when loads were added or removed. A change in load resulted in a change in joint angle (equilibrium point) which was predictable as the behavior of a non-linear spring (Asatryan & Fel'dman, 1965).

The question arises as to how such a spring might be controlled in order to produce different steady state positions. According to Fel'dman (1966) (see also Houk, 1978), this can be accomplished by adjusting certain parameters--'tuning' the spring--prior to movement. In this account, the nervous system sets the values of resting length, λ, by adjusting the length-tension relationships of the muscles involved. If the length of the muscle, λ, varies from the resting length, "voluntary" movement takes place. If $\chi > \lambda$ an active tension develops in the muscle and if $\chi < \lambda$ the muscle is relaxed. The invariant character of the muscle is, therefore, the dependence of tension on length for any fixed value of λ. Thus the only static parameter which need be set for voluntary movement in Fel'dman's model is resting length--namely, the length of the muscles for which differences in tension sum to zero. On the other hand, kinematic changes in rate, acceleration, and periodicity in the joint-muscle collective are brought about by altering the dynamic parameters of stiffness and damping.

Recent data fit this perspective rather well, at least on a posteriori grounds. For example, Bizzi and his colleagues (Bizzi, Dev, Morasso, & Polit, 1978; Polit & Bizzi, 1978) have shown for both head and arm movements that normal and rhizotomized monkeys can accurately achieve learned target positions even when constant and brief load perturbations were applied during the movement trajectory. They argue that the controlled variable must be an equilibrium point specified by the motor program in terms of the length-tension relationships in agonist and antagonist muscles. Similarly, a consistent outcome in human experiments has been the superior accuracy of attaining final position over amplitude from variable starting positions; a finding that extends to functionally deafferented subjects (Kelso, 1977) as well as patients in whom positional detectors in the joint capsule have been surgically removed.[2] Terminal location may be viewed as an equilibrium point specified by the tuned parameters of the spring: it is thus impervious to unforeseen changes in initial starting position. Amplitude production, on the other hand, involves a change in the equilibrium point as a function of variable initial conditions, and hence a re-parameterization of spring parameters.

In the present experiments, we set out to determine--on an a priori basis--whether any of the observed kinematic characteristics that arise in localization violate the mass-spring model. Specifically, our tack was to introduce sudden and unexpected torque loads--which acted to drive the limb (in this case the index finger) in the opposite direction--and observe consequent effects on localization. Unlike numerous other studies (see Desmedt, 1978, for a review) we were not particularly concerned with evaluating the various reflex responses to changed loading conditions. Rather we wished to elucidate the effects of changing dynamic parameters and consequent kinematic variation on the attainment of a specified equilibrium position. In Experiment 1 we show that the equilibrium position is accurately attained despite on-line perturbations. In Experiment 2 we rule out possible alterna-

tive accounts by replicating this result in functionally deafferented individuals.

EXPERIMENT 1

Method

The subjects were 12 (8 women, 4 men) right-handed graduate and undergraduate students who volunteered for the experiment and were not paid for their services. The apparatus consisted of a finger positioning device (see Figure 1) and the associated programming electronics. The movements allowed by the positioning device were flexion and extension of the index finger about the metacarpophalangeal joint. The distal end of the moving finger was fitted with a plastic collar that slipped into an open-ended cylindrical support. The support revolved about the metacarpophalangeal joint and prevented movement of the distal joints of the finger. Attached to the end of the support was a pointer that moved over a protractor graduated in degrees. The device was also equipped with padded adjustable clamps with which to secure the subject's wrist, hand and remaining fingers and thumb during the movements.

The electronics control package supported the programming of torque motor output with respect to the movement of the finger. Finger movements could be loaded by a system of gears that were driven by the motor producing a maximum of 81.6 oz/in of torque about the joint. A control was available that varied the amount of resistance that could be applied before the perturbation (pre-perturbation) and after the perturbation (post-perturbation). A second potentiometer enabled control of the amount of torque applied during the perturbation. The location (angle) at which the perturbation was triggered, as well as its duration, could be controlled directly from the electronic panel. A potentiometer mounted over the axis of motion provided information regarding the position and velocity of movement for recording purposes.

Electromyographic potentials were recorded with Beckman silver--silver chloride disc type electrodes and amplified and recorded on a Beckman 5010 Polygraph System.

Procedure

The skin above the right extensor digitorum and over the lateral aspect of the olecranon process (ground electrode) was prepared for electrode placement. In some cases the skin over the flexor digitorum superficialis was also prepared. The interelectrode resistances were 5K ohms or lower.

The subjects were blindfolded and seated in a dental chair such that the right arm and index finger could be comfortably but securely arranged in the positioning apparatus. The procedures for movement during acquisition and perturbation trials were then described for the subject and any questions answered. Through the experiment the subject's task--on the commands "ready" and "move"--was simply to move rapidly to the designated position. At that point the subject was to say "there" following which the experimenter returned the finger to the starting position. The experiment proceeded in two parts. The first, acquisition trials, consisted of 30 extension movements to a to-be-

Table 1

Means and standard deviations of absolute, constant, and variable error (in degrees) and movement time (in msec) for acquisition, perturbed and non-perturbed movements, Experiment 1.

MEANS		Acquisition[a]	Non-Perturbed[b]	Perturbed[b]	Perturbed[c]		
					Short	Medium	Long
Absolute	M	2.81	4.12	5.61	6.08	6.17	4.58
Error	SD	1.21	2.44	2.27	3.96	3.74	2.74
Constant	M	-0.15	-0.18	1.37	0.58	1.61	1.97
Error	SD	1.88	3.95	5.33	7.33	6.20	4.89
Variable	M	3.14	4.11	4.54	3.01	4.10	3.16
Error	SD	1.30	2.08	1.50	1.50	3.19	1.89
Movement	M		194	346	302	325	410
Time	SD		43	62	20	30	35

[a]Means of last nine acquisition trials
[b]Means of nine trials
[c]Means of three trials

learned target position (50 degree movement from the starting position that remained constant at 20 degree flexion). Verbal knowledge of results (KR) was given in qualitative (overshoot, undershoot, hit) and quantitative (number of degrees) terms. Movement errors were recorded in degrees of angular displacement from the target position. Following the acquisition trials there were 18 test trials (without KR) of which half were perturbed by the programmable torque motor. The locations of the perturbation were designated as short (applied after 10 degrees of movement from the starting position), medium (after 25 degrees of movement), or long (after 40 degrees of movement). There were three trials at each of the three perturbation locations, and these were randomly ordered amongst the 18 test trials. The subjects were informed that on some of the trials a perturbation would occur, and that they should attempt to move through it and arrive at the learned location. The duration of the application of the perturbation was set at 100 msec throughout the experiment. Free movement was allowed in non-perturbed trials. During perturbed trials the movement was essentially free before and after the 100 msec torque duration. Extensor and flexor muscle activity, position, and velocity of movement were recorded simultaneously on the polygraph. Acquisition trials were recorded at 10 mm sec^{-1}, and test trials at 50 mm sec^{-1}.

Results

For error scores, deviations from the target position were recorded. By convention an undershoot was signed negative (-) and an overshoot was signed positive (+). Absolute error (unsigned), constant error (signed) and variable error (standard deviation around mean constant error) were used for the purposes of analysis.

Acquisition trials. Acquisition trials were divided into six blocks of five trials and the plot is presented in Figure 2. As can be seen in the figure, there were improvements in performance as reflected in significant differences between trial blocks 1 and 6, $t(11) = 3.62$, $p < .01$ and 3.99, $p < .01$ for absolute and variable error respectively. An analysis of constant error failed to reveal significance, $p > .05$.

Test trials. To test for the principle of equifinality in our subjects, the nine non-perturbed trials were compared to the nine perturbed trials. Means were calculated for each subject (see Table 1) and served for paired t-tests. The contrast between perturbed and non-perturbed trials failed significance for absolute error, ($t(11) = 2.01$ $p > .05$). Constant and variable error comparisons revealed similar results; $t(11) = 1.51$, $p > .05$, and $t(11) = .74$, $p > .05$, respectively. Examination of the raw absolute error data revealed that nine of the 12 subjects showed little or no decrement in performance as a result of the perturbations. Perturbed trials were further subdivided according to the locus of perturbation (short, medium, or long) and analyzed in a one-way analysis of variance. No main effects were found for locus of perturbation, $Fs(2,33) = .77$, $p > .05$, .16, $p > .05$ and .79, $p > .05$ for absolute, constant and variable error respectively.

For a physical mass-spring with constant stiffness and damping parameters certain invariant kinematic details will emerge. Velocity and periodicity, for example, are constants as is the overall movement time. Mean movement time data calculated from the first overt sign of movement in the potentiome-

ter output to the point of target attainment are also presented in Table 1.

To test whether velocity was constant from trial to trial in this learned movement, velocity was computed from the linear part of the slope of the displacement curves for non-perturbed trials and averaged for all subjects (mean = 350 deg sec^{-1}, SD = 86 deg sec^{-1}). These values represent a mean variability of about 25% and none of the subjects showed less than 16% variability. Thus in a learned positioning task without external perturbations, it is clearly seen that the emergent kinematic variability in movement time and velocity indicates that one or both of the underlying dynamic parameters of stiffness or damping undergo change. To test this finding further we investigated the oscillations of the finger around the equilibrium point. A spring system with constant stiffness and damping parameters will always produce one of three kinds of oscillation: light, critical, or heavy damping (Volterra & Zachmanoglou, 1965). A lightly damped system is defined as one in which the mass of the spring passes through the equilibrium point at least once before reaching steady state or equilibrium point. A heavily damped system has the characteristic that the mass does not pass through the equilibrium point and only reaches equilibrium at infinity. The critically damped system is one where the damping has the least value that will produce aperiodic motion; that is, the value at which the spring moves quickly to the equilibrium point without ever passing through it. Actual displacement curves of each of these forms of damping are presented in Figure 3. The above criteria were used to determine qualitatively whether a subject displayed one or more of these movement patterns. On examination of the raw data it was observed that of the 12 subjects, nine demonstrated both critical and light damping characteristics and none showed heavy damping. There was a tendency towards critical damping in non-perturbed trials (76% of all trials critically damped) while in perturbed trials there was a slight tendency for light damping (54% lightly damped). Locus of perturbation had no obvious effect on oscillatory activity (53%, 64% and 44% critically damped for perturbed short, medium and long trials respectively).

Discussion

The finding that equally accurate performance was obtained in both perturbed and non-perturbed trials strongly supports the equifinality property that is characteristic of vibratory systems. Although comparisons are tenuous, our results appear even more favorable for the concept than those obtained on arm movements in monkeys where the errors are quite large (Polit & Bizzi, 1978) (see Figure 2). In addition, this is the first time to our knowledge that equifinality in the face of unpredictable perturbations has been observed in human subjects. Perhaps surprisingly, the kinematic variability in velocity, movement time and oscillations points also to variability in at least one of the dynamic parameters of stiffness and damping. We might have expected that in a learned motor activity, the central nervous system would maintain these parameters constant from trial to trial. However, it may be argued that since the task demanded only target attainment, the movement patterns by which this goal was achieved were relatively unimportant. Thus the control system could feasibly afford several variations in the specification of parameters for achieving the equilibrium position. That there was a tendency towards greater amplitude of oscillation in perturbed trials may be viewed as a result of injecting more energy into a system with constant

damping, an interpretation that is congruent with the dynamic response characteristics of a mass-spring system.

Clearly, the data obtained in Experiment 1 can more than adequately be explained by a mass-spring model of localization. However, the findings fail to rule out an alternative hypothesis based on a closed-loop conceptualization of control. Thus our results could be accounted for by supposing that central commands are modified during movement execution by the action of fast acting peripheral feedback loops (Cooke & Eastman, 1977; Evarts & Granit, 1976; Houk, 1978), or, given sufficient time, through resetting the central commands following perturbation. Although the latter possibility is unlikely, due to the ballistic nature of the movements employed (see Table 1) we do not rule out the role of afferent information in re-parameterizing the system following movement termination. A major prediction with respect to the spring model, however, is that a continuous readout of proprioceptive information is not a necessary condition for the achievement of the equilibrium position. The second experiment was designed to examine this issue by injecting load perturbations during the localization movements of individuals who had joint and cutaneous information removed using the wrist cuff technique (Goodwin, McCloskey, & Matthews, 1972; Kelso, 1977; Merton, 1964). The advantage of this procedure is that muscle function is preserved in the long finger flexors and extensors that lie high in the forearm, while sensory inputs to the hand itself are effectively eliminated.

EXPERIMENT 2

Method

Six volunteers from the student and faculty population served as subjects. None of the subjects had been involved in the first experiment. All were informed of the sensations involved in the wrist cuff procedure, and only those who felt they could handle the situation participated.

The apparatus used in this experiment was identical to that employed in Experiment 1.

Procedure

We felt it necessary to replicate entirely the results of the first experiment. Thus, the first part of Experiment 2 followed the earlier procedure exactly. After the 30 acquisition trials with knowledge of results available, nine perturbed and nine non-perturbed trials (together designated pre-cuff trials) were given in randomized order with error information withdrawn. On completion of this aspect of the study, subjects were removed from the apparatus and the wrist cuff applied and inflated as discussed in detail elsewhere (Kelso, 1977). The subject's arm and hand were replaced in the apparatus in the same posture as before. Sensory testing was conducted at one minute intervals 5 to 10 min after wrist cuff application in order to let the subjective experience of "pins and needles" subside. Subjects were instructed to label the fingers one through five (from thumb to little finger) and to specify the sensation tested (e.g., "Touch four," "Movement two"). Tactile sensation was tested using a cotton covered stick applied to the skin. Movement-position sense was similarly tested by the experimenter moving the

Table 2

Means and standard deviations of absolute, constant, and variable error (in degrees) and movement time (in msec) for acquisition, perturbed and non-perturbed movements, Experiment 2.

MEANS		Acquisition[a]	Non-Perturbed[b]	Perturbed[b] Total	Perturbed[c] Short	Medium	Long
Absolute	M	2.33	5.61	5.21	6.22	3.84	5.56
Error	SD	0.80	3.42	2.48	4.46	1.62	2.18
Constant	M	0.33	-4.50	-3.72	-5.11	-1.95	-3.67
Error	SD	1.89	4.19	3.80	5.90	3.23	3.74
Variable	M	2.52	3.35	4.36	3.31	2.87	4.07
Error	SD	0.55	1.06	0.83	1.22	0.75	2.39
Movement	M		192	352	317	335	396
Time	SD		39	53	28	40	47

[a]Means of last nine acquisition trials
[b]Means of nine trials
[c]Means of three trials

Table 3

Means and standard deviations of absolute, constant, and variable errors (in degrees) and movement times (in msec) for perturbed and non-perturbed movements under wrist-cuff conditions, Experiment 2.

MEANS		Non-Perturbed[a]	Perturbed[a]	Perturbed[b]		
			Total	Short	Medium	Long
Absolute Error	M	13.52	11.66	13.56	10.28	9.58
	SD	7.87	5.88	9.66	6.14	3.56
Constant Error	M	4.89	0.00	-3.55	4.61	4.25
	SD	15.74	12.78	16.03	11.09	8.67
Variable Error	M	6.25	8.36	8.46	5.07	6.53
	SD	2.28	2.57	3.94	1.99	3.53
Movement Time	M	417	454	418	451	485
	SD	90	79	52	50	39

[a]Means of nine trials
[b]Means of three trials

digits in specific directions at variable rates. The order of stimulus presentation was random for both modalities to further insure discriminating responses on the part of the subject. Occasionally catch trials were included in which no stimulation was given in order to reduce guessing by the subject. When the subject was no longer capable of reporting touch or localizing position and movement in the digit being tested on two consecutive occasions, the experimenter defined this point as the respective cut-off point for that digit. When all digits succumbed to this criterion, the final endpoint was assumed and its time of occurrence recorded. Following the establishment of sensory cut-off3 a further 18 trials were given to the subject, half of which were perturbed at three different loci. These trials (designated cuff trials) were yoked to the pre-cuff trials so that the subject performed them in the same randomized order. Again, these were performed in the absence of knowledge of results. On completion of these movements the wrist cuff was deflated and the subject remained seated for a mandatory recovery period of 10 min before leaving the laboratory.

Results

Acquisition trials. Almost identical results to those obtained in Experiment 1 were obtained for acquisition trials (see Figure 2). The comparison of trial blocks 1 and 6 for both absolute and variable error were significant, $t(5) = 10.9$, $p < .01$ and 5.03, $p < .01$ respectively.

Pre-cuff test trials. These trials were analyzed as in Experiment 1. Absolute, constant and variable error means are presented in Table 2. No significant differences were found in the pairwise comparisons of perturbed and non-perturbed trials for absolute and constant error, $t(5) = .86$, $p > .05$ and 1.52, $p > .05$, respectively. Significantly larger variable error was found in the perturbed trials, $t(5) = 4.03$, $p < .01$. However, as can be seen in Table 2, this difference was small, in the order of 1 deg. Examination of the absolute error means revealed that all six subjects showed little or no decrement in performance on perturbed trials. Locus of perturbations was not found to be a significant factor as indicated by analysis of variance, $F(2,15) = 1.00$, $p > .05$, .76, $p > .05$, and .84, $p > .05$ for absolute, constant and variable error respectively. Mean movement time data are also presented in Table 2 and mirror those of the previous study. Results of velocity computations were similar to those of Experiment 1 (mean = 346 deg sec^{-1}, SD = 67 deg sec^{-1}), with the exception that the mean variability in this case was slightly lower (20%).

Four of the six subjects showed oscillations in movement pattern indicative of both light and critical damping. The effects of perturbation were even more pronounced in this experiment. While 70% of non-perturbed trials revealed critical damping, only 30% of perturbed trials were critically damped. Again, locus of perturbation had no obvious effects on oscillatory behavior with critical damping in 28%, 28% and 33% of perturbed short, medium and long trials respectively.

Cuff trials. Mean error scores for subjects performing under wrist cuff conditions are presented in Table 3. A comparison of non-perturbed and perturbed trials revealed no significant differences for absolute error, $t(5) = .19$, $p > .05$. Analysis of constant, $t(5) = 3.33$, $p < .05$ and variable

error, $t(5) = 3.88$, $p < .05$, however, indicated significant differences. The mean constant error for non-perturbed trials was larger and more positive than that for perturbed trials, while the variability in the perturbed trials was greater than in non-perturbed trials. It may be noted, however, that the differences are very modest indeed compared to the boundary conditions set by Polit and Bizzi (1978) for accurate arm movements in monkeys (in the order of 12 to 15 deg).

Locus of perturbation was found not to be a significant factor as indicated by analysis of variance, $F(2,15) = .46$, $p > .05$, .74, $p > .05$ and 1.21, $p > .05$ for absolute, constant and variable error respectively.

Movement time data are also presented in Table 3. Interestingly, there was an overall increase in movement time in cuff trials and this was accompanied by changes in both velocity and oscillation patterns. The mean velocity for wrist cuff trials was less than that of pre-cuff trials (mean = 260 deg sec^{-1}, SD = 80 deg sec^{-1}, representing a mean variability of 31%). Without exception, the displacement curves for all subjects under both perturbed and non-perturbed conditions showed a critical damping pattern only.

It is not legitimate to compare accuracy scores from pre-cuff to cuff trials due to the substantial time lapse that was necessary for the nerve block to take effect (between 1 and 1.5 hr in all subjects). Thus the modest increase in error is likely accounted for by the time delay combined with the absence of knowledge of results regarding performance. Of course, to completely discount the possibility of proprioceptive influences on target accuracy in perturbed and non-perturbed cuff trials is not possible within the present experimental paradigm (see General Discussion).

Electromyographic data. Qualitative differences in EMG activity were examined in pre-cuff and cuff trials. Examples are given in Figure 4 along with accompanying displacement records illustrating perturbed and non-perturbed movements. As shown in figure 4(b) there is an increase in agonist EMG activity following perturbation onset due presumably to proprioceptive stimulation and consequent initiation of fast acting reflex loops (e.g., Evarts, 1973; Marsden, Merton, & Morton, 1976). In cuff performance, however, electrical activity was constant throughout the movement and signs of stretch reflex activity were largely absent [Figure 4(c) and (d)]. A notable observation was that the activity of the antagonist muscle (flexor digitorum superficialis) was close to baseline during pre-cuff trials but highly active during and after target localization in cuff conditions.

Discussion

There were two principal outcomes of this experiment. First it replicated in full the major results of our first experiment. Equifinality was observed under normal localization conditions and when movements were unexpectedly perturbed by suddenly applying a torque load that drove the finger in the opposite direction (see Figure 4). Second, the equifinality characteristic was present even when the subject was functionally deafferented, a finding that lends converging support to the mass-spring model.

An additional and interesting result was that movements in the wrist cuff condition were slower and movement patterns more consistent than in pre-cuff trials. Bizzi et al. (1978) also observed this differential effect in head movements between animals with and without intact proprioception, a finding they attribute to the fact that only the mechanical properties of the neck musculature (active and passive) remain for control purposes. In our study, an observed increase in EMG activity in flexor and extensor muscle groups under cuff conditions was combined with a perceived increase in effort in all subjects. We might expect that an increase in the conjoint activity of flexor and extensor muscles will have consequent stiffening effects on the metacarpophalangeal joint. Given parallel increases in frictional forces and muscle stiffness, the system will convert from lightly damped to critically damped (Volterra & Zachmanoglou, 1965), thus suggesting a reason why all our subjects showed critical damping in cuff trials.[4] It is also possible that the modest loss in accuracy observed under cuff conditions occurs as a result of the disruption of stiffness ratios in antagonist muscles that normally hold for particular equilibrium positions.

In sum, the results of Experiment 2 suggest that proprioceptive information (as far as can be determined in a human preparation) is not a requirement for the accurate attainment of equilibrium position. This conclusion generally parallels very recent work showing that the neural reflex component has a relatively small contribution (10 to 30%) in counteracting applied disturbances during movement localization. In contrast, a significantly larger portion (60%) of this process is provided by mechanical properties (inertial, viscous, and elastic) of the neck musculature (Bizzi et al., 1978). Moreover, none of our findings refute the claim that the kinematic characteristics of movement observed in this experiment (or in Experiment 1) violate any of the dynamic laws pertaining to a mass-spring system.

GENERAL DISCUSSION

The present experiments were designed to test the efficacy of a mode of control that takes advantage of the properties of a particular type of vibratory system, namely the mass-spring. A most important characteristic of a mass-spring system is its equifinality (von Bertalanffy, 1973), which emerges as the predominant feature in our data. Thus equifinality endured despite unexpected and abrupt load disturbances (Experiment 1), functional deafferentation and both in conjunction (Experiment 2). These results fully complement our earlier work (Kelso, 1977) and, corroborated by recent neurophysiological data (Bizzi et al., 1978; Houk, 1978), provide a broad empirical basis for the mass-spring model.

There are some grounds for caution, however. Although the pressure cuff technique drastically reduces joint and cutaneous information, it is possible that muscle proprioception plays a role at some level in the achievement of final position. Although we envisage spring parameters to be set prior to movement, we cannot exclude muscle and tendon reception as playing a regulatory function. Regardless of whether this is the case or not, our basic tenet--borne out by the data and consonant with the mass-spring model--is that it is dynamic variables that are regulated.

It might further be argued that the significant increase in target acquisition time following perturbations observed in both experiments is a consequence of "reaction-time" movements that are proprioceptively based (e.g., Houk, 1978, for review). While we cannot completely rule out such a view we should reemphasize that the duration of the perturbation was 100 msec. Inspection of Table 1 shows that unless such reaction time components are embedded in the perturbation duration, the response at the short perturbation locus for example, would be around 8 msec, which is well outside the boundaries of proprioceptive reaction time. Perhaps a more parsimonious explanation exists in terms of the amount of force produced by the limb muscles at different stages during the trajectory. For instance, early in the movement larger forces are necessary to overcome inertia and to accelerate the limb. Consequently, the effect of a perturbation (a force vector operating in the opposite direction) will be less than in the case where the limb is at constant velocity or decelerating. Thus the locus of the perturbation (whose characteristics are constant for all trials) in the movement trajectory will determine the amount of deflection of the limb and the time to return the limb to the point at which it was perturbed. The systematic effects of locus of perturbation on target acquisition time (see Tables 1 and 3), render this explanation as a viable--and testable--alternative to a "reaction time" account.

Given the foregoing caveats, let us now turn to the broader implications of our findings. The mass-spring model of localization promoted here represents a significant departure point from other conceptualizations of control. It is fairly common, for example, for researchers to speak of central motor programs for movement amplitude and duration (Brooks, 1974; Taub et al., 1975), or to view the achievement of final position (location) as the continually sensed position of the limb in reference to a perceptual referent or spatial coordinate system (Russell, 1976). Although these modes of control are conceptually distinct--and may refer to different types of movement (e.g., ballistic versus slow [Desmedt & Godeaux, 1978]), they nevertheless envisage the controlled variable as a kinematic prescription. In contrast, the behavior of a mass-spring system, while describable in kinematic terms such as displacement, rate, and frequency, is totally determined by dynamic properties such as mass, viscosity, and stiffness. Although it is difficult to conclusively account for the kinematic variability in our data in terms of a singular dynamic parameter, there is good reason to believe that stiffness is the regulated entity, while damping is a constant of the viscous properties of the joint and muscle groups involved (Fel'dman, 1966; Houk, 1978; Nichols & Houk, 1976).

What theoretical advantages might a mass-spring account of localization provide? An adequate answer to this question requires us to briefly address two fundamental problems of motor organization that are often overlooked in both behavioral and neurophysiological investigations of motor mechanisms. The first of these is the issue of <u>functional non-univocality</u> (Bernstein, 1967) or context-conditioned variability (Turvey, Shaw, & Mace, 1978); simply stated, there is no invariant relationship between centrally generated signals and movement outcomes. Movements cannot be direct reflections of neural events because muscular and non-muscular (reactive) forces must be taken into account. Similarly, at a cellular level, direct monosynaptic control of alphamotoneurons is the exception rather than the rule in neural regulation of

movement. Whether a motoneuron fires or not is contingent on the influences of supra-, inter-, and intra-segmental neurons whose status varies from one instant to the next (Evarts, Bizzi, Burke, DeLong, & Thach, 1971). The point is that the effects of descending influences are continually modulated by virtue of the active state of the segmental machinery. Thus we cannot ignore the contextual background against which supraspinal signals are realized.

Although it is possible, in theory, to solve the problem of context-conditioned variability by making available detailed information about the current states of muscles and joints, such an account fails to address the second problem: namely, how the degrees of freedom of the motor apparatus are regulated.5 One solution towards resolving this dilemma (Bernstein, 1967; Gelfand, Gurfinkel, Fomin, & Tsetlin, 1971; Greene, 1972; Turvey, 1977) is to claim that skeletomuscular variables are partitioned into collectives where the variables within a collective change relatedly and autonomously. This results in a reduction of the degrees of freedom, since higher brain centers now have only to activate or 'tune' lower-level, functional groups of muscles.

That such functional units or synergies (Boylls, 1975) constitute the significant units of control is revealed in a broad range of activities--from the regulation of animal gait (Grillner, 1975; Shik & Orlovskii, 1965), and the maintenance of vertical posture (Nashner & Grimm, 1978) to aiming the arm at a target (Arutyunyan, Gurfinkel, & Mirskii, 1968) and coordinating the upper limbs to perform different tasks (Kelso, Southard, & Goodman, 1979). Similarly, 'tuning' is demonstrated by experiments that sample the state of the neuromuscular system just prior to movement. Thus it can be shown that augmentation of the H-reflex in the gastrocnemius muscle occurs when it is compatible with the intended movement (plantar flexion) and depressed when it is not (Gottlieb, Agarwal, & Stark, 1973; Kots, 1977). Functional groupings of muscles arise then, presumably as a result of such tuning or biasing in the interneuronal pools of the spinal cord (Gelfand et al., 1971; Greene, 1972).

The mass-spring model is fully compatible with the style of control that we have briefly elaborated here. The spring system in this case represents a functional collective of muscles (Greene, 1972; Turvey, 1977), and its dynamic parameterization is reflective of tuning. It thus provides at least a partial solution to some of the problems that confront investigations of movement production possessing the following advantages: First, the desired position of the limb may be specified independent of initial conditions (equifinality). Thus, what constitutes a context-conditioned variability problem in traditional views of motor control is not a concern for the mass-spring proposition. Second, the desired behavior of the limb is closely approximated by the laws of mechanics. Given the existence of constant damping, the limb pursues its own trajectory in a stable manner. Both these advantages eliminate the requirement for continuous regulation via feedback. Third, speed can be controlled by specifying stiffness, which, as revealed in elementary mechanics as well as in the present experiments, may be regulated independently of resting length (equilibrium point). Thus, an invariant endpoint may be the outcome of varying combinations of kinematic properties. Fourth, repetitive movements can be produced by the natural oscillations of a spring system; all that needs to be specified--and then only once--is the equilibrium point. Instead of controlling time of arrival at each terminal position of the swing (a kinematic computation) timing could be regulated by

altering stiffness (the spring constant). A fifth advantage has been suggested by Greene (personal communication) and is demonstrated in the following example: Consider the task of directing a finger to a desired position in space. Spring parameters of the trunk, shoulder, elbow, and wrist could be temporarily coupled so that when the equilibrium point of the entire linkage is reached, the finger is in the correct position.6 If for some reason the location is not reached, then the position error of the finger alone--not the n-tuple errors of all the degrees of freedom--can be used to regulate the equilibrium point of the coupled system. This is an important point for it reconciles to some degree the problem that some theorists have had (Fowler & Turvey, 1978) with formulations that have knowledge of results regulate motor learning (Adams, 1971). If the skeletomuscular variables are constrained to act as a unit, error information will be relevant to that unit alone and not the singular degrees of freedom that constitute it.

Finally, a neuromuscular system with the dynamic properties of a spring complements the dynamic nature of the environment. For example, in a recent paper, McMahon and Greene (1978) have demonstrated that running performance shows impressive improvement when the compliance of the track surface is closely matched to the spring parameters of the runner. Also, as Houk (1978) points out, when we encounter an immovable obstacle during movement, it is more beneficial for the musculature to yield, rather than to needlessly regulate joint position since this would require a complete readjustment of other body parts. In summary, we feel there is much to commend a mass-spring account especially in terms of economy of neural control. A number of challenging questions remain, however, particularly those that address the physiological nature of the 'tuning' process and how the present analysis may be extended to continuous movements.

SUMMARY AND CONCLUSIONS

1. Two experiments were conducted to determine the efficacy of a vibratory systems analysis of finger localization on humans. In both experiments the subject's task was to move the finger rapidly to a previously learned target location. The second experiment also included a condition in which subjects were functionally deafferented by means of a child's sphygmomanometer cuff inflated around the wrist. In both experiments the finger was displaced on 50% of the test trials by brief torque loads (perturbations) injected at unpredictable points during the movement.

2. It was found that in both the normal and functionally deafferented subjects, perturbations did not affect accuracy in achieving final position, a finding that supports the equifinality characteristic of a mass-spring. That is, a mass-spring will always reach an invariant final position or equilibrium point no matter how perturbed. In addition, it was found that none of the kinematic analyses computed (velocity, movement time, oscillation) violated the dynamic behavior of the mass spring model.

3. A control system that takes advantage of the visco-elastic properties of the muscles and joints reduces the control problem for the brain in that once underway, movements to endpoints need no moment-to-moment intervention by central mechanisms. Furthermore, initial conditions such as starting position are of little concern in this mode of control: That

is, as long as an endpoint is specified, initial starting position and brief perturbations have no consequence on the endpoint reached.

4. Theoretically this finding is of some significance since it reduces two problems for control of movement by the brain: degrees of freedom and context-conditioned variability.

REFERENCES

Adams, J. A. A closed loop theory of motor learning. Journal of Motor Behavior, 1971, 3, 111-150.

Adams, J. A. Feedback theory of how joint receptors regulate the timing and positioning of a limb. Psychological Review, 1977, 84, 504-523.

Arutyunyan, G. A., Gurfinkel, V. S., & Mirskii, M. L. Investigation of aiming at a target. Biophysics, 1968, 13, 642-645.

Asatryan, D. G., & Fel'dman, A. G. Functional tuning of the nervous system with control of movement or maintenance of a steady posture - I. Mechanographic analysis on the work of the joint on execution of a postural task. Biophysics, 1965, 10, 925-935.

Bernstein, N. A. On the construction of movements, Monograph (in Russian), Moscow, 1947.

Bernstein, N. A. The coordination and regulation of movements. London: Pergamon Press, 1967.

Bizzi, E., Dev, P., Morasso, P., & Polit, A. Effects of load disturbances during centrally initiated movements. Journal of Neurophysiology, 1978, 41, 542-556.

Bizzi, E., & Polit, A. Processes controlling visually evoked movements. Neuropsychologia, in press.

Boylls, C. C. A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. COINS Technical Report (Department of Computer and Information Science, University of Massachusetts, Amherst), 1975, 76-1.

Brooks, V. B. Some examples of programmed limb movements. Brain Research, 1974, 71, 299-308.

Cooke, J. D., & Eastman, M. J. Long-loop reflexes in the tranquilized monkey. Experimental Brain Research, 1977, 27, 491-500.

Desmedt, J. E. (Ed.) Cerebral motor control in man: Long loop mechanisms. Basel: Karger, 1978.

Desmedt, J. E., & Godeaux, E. Ballistic skilled movements: Load compensation and patterning of the motor commands. In J. E. Desmedt (Ed.), Cerebral motor control in man: Long loop mechanisms. Basel: Karger, 1978.

Evarts, E. V. Feedback and corollary discharge: A merging of the concepts. Neurosciences Research Program Bulletin, 1971, 9, 86-112.

Evarts, E. V. Motor cortex reflexes associated with learned movement. Science, 1973, 179, 501-503.

Evarts, E. V., Bizzi, E., Burke, R. E., DeLong, M., & Thach, M. T. (Eds.). Central control of movement. Neurosciences Research Program Bulletin, 1971, 9, 1-171.

Evarts, E. V., & Granit, R. Relations of reflexes and intended movements. Progress in Brain Research, 1976, 44, 1-14.

Fel'dman, A. G. Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic

analysis of execution by man of the simplest motor tasks. Biophysics, 1966, 11, 766-775.

Fowler, C. A., & Turvey, M. T. Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic, 1978.

Gelfand, I. M., Gurfinkel, V. S., Fomin, S. V., & Tsetlin, M. L. Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

Goodwin, G. M., McCloskey, D. I., & Matthews, P. B. C. The contribution of muscle afferents to kinesthesia shown by vibration induced illusions of movement and by the effects of paralyzing joint afferents. Brain, 1972, 95, 705-748.

Gottlieb, G. L., Agarwal, G. C., & Stark, L. Interaction between voluntary and postural mechanisms of the human motor system. Journal of Neurophysiology, 1973, 33, 365-381.

Greene, P. H. Problems of organization of motor systems. In R. Rosen and F. Snell (Eds.), Progress in theoretical biology (Vol. 2). New York: Academic, 1972.

Grillner, S. Locomotion in vertebrates: Central mechanisms and reflex interaction. Physiological Review, 1975, 55, 247-304.

Houk, J. D. Participation of reflex mechanisms and reaction time processes in compensatory adjustments to mechanical disturbances. In J. E. Desmedt (Ed.), Cerebral motor control in man: Long loop mechanisms. Basel: Karger, 1978.

Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 529-543.

Kelso, J. A. S., Holt, K. G., & Flatt, A. E. Towards a theoretical reassessment of the role of proprioception in the perception and control of human movement. Haskins Laboratories Status Report on Speech Research, 1979, SR-58, 1-12.

Kelso, J. A. S., Southard, D., & Goodman, D. On the nature of human interlimb coordination. Science, 1979, 203, 1029-1031.

Kelso, J. A. S., & Stelmach, G. E. Central and peripheral mechanisms in movement. In G. E. Stelmach (Ed.), Motor control: Issues and trends. New York: Academic, 1976.

Kots, Ya. M. The organization of voluntary movements. New York: Plenum, 1977.

Marsden, C. D., Merton, P. A., & Morton, H. B. Stretch reflexes and servoactions in a variety of human muscles. Journal of Physiology, London, 1976, 259, 531-560.

McMahon, T. A., & Greene, P. R. Fast running tracks. Scientific American, 1978, 240, 148-163.

Merton, P. A. Human position sense and sense of effort. Symposium of the Society of Experimental Biology, 1964, 18, 387-400.

Nashner, L. M., & Grimm, R. J. Analysis of multiloop dyscontrols in standing cerebellar patients. In J. E. Desmedt (Ed.), Cerebral motor control in man: Long loop mechanism. Basel: Karger, 1978.

Nichols, T. R., & Houk, J. C. The improvement of linearity and the regulation of stiffness that results from the actions of the stretch reflex. Journal of Neurophysiology, 1976, 39, 119-142.

Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys.

Science, 1978, 201, 1235-1237.
Russell, D. G. Spatial location cues and movement production. In G. E. Stelmach (Ed.), Motor control: Issues and trends. New York: Academic, 1976.
Schmidt, R. The schema as a solution to some persistent problems in motor learning theory. In G. E. Stelmach (Ed.), Motor control: Issues and trends. New York: Academic, 1976.
Shik, H. L., & Orlovksii, G. N. Coordination of the limbs during running of the dog. Biophysics, 1965, 10, 1148-1159.
Stelmach, G. E., Kelso, J. A. S., & Wallace, S. Preselection in short-term motor memory. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 745-755.
Taub, E., Goldberg, I. A., & Taub, P. Deafferentation in monkeys: Pointing at a target without visual feedback. Experimental Neurology, 1975, 46, 178-186.
Tomovic, R., & Bellman, R. A systems approach to muscle control. Mathematical Bioscience, 1970, 8, 265-277.
Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, N.J.: Erlbaum, 1977.
Turvey, M. T., Shaw, R., & Mace, W. Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Erlbaum, 1978.
Volterra, E., & Zachmanoglou, E. C. Dynamics of vibrations. Columbus, Ohio: Merrill, 1965.
von Bertalanffy, L. General systems theory. London: Penguin Press, 1973.

FOOTNOTES

[1]Peter Greene (personal communication) alerted us to this fact for which we are grateful. The translated version of Bernstein's work (Bernstein, 1967) does not include a discussion of the mass-spring model.

[2]Data collected on patients with metacarpophalangeal joints removed and replaced with silastic inserts (see Kelso, Holt, & Flatt, 1979). No deficits in finger positioning capabilities were found.

[3]Three independent sources of evidence speak to the viability of the wrist cuff technique as a tool in reducing proprioceptive sensations. First, passive displacements of the metacarpophalangeal joint up to an estimated 90 degrees per sec were undetected. Second, subjects when instructed to produce a movement but prevented from doing so, consistently perceived that they had executed the movement. If muscle afferent information were capable of accessing consciousness, this would have been an unlikely finding. Third, it has been consistently verified that the loss of background facilitation from joint and cutaneous sources using this procedure eliminates stretch reflex function (e.g., Marsden, Merton, & Morton, 1972; Merton, 1974).

[4]This outcome is evident on consideration of the equation of motion for a freely damped mass-spring system:

$$m\ddot{x} + c\dot{x} + kx = 0$$

where m is the mass, c is the linear damping constant and k is stiffness. It can be shown that where $c^2 = 4$ mk critical damping occurs and where $c^2 < 4$ mk light damping occurs. Thus, the relationship between spring constant (k), damping constant (c) and mass (m) determines the system's oscillation.

[5]If we consider the degrees of freedom problem at the level of muscles alone, Tomovic and Bellman (1970) have estimated the total number in the human body to be 792. For the brain to individually regulate these is surely an insurmountable task.

[6]In fact, this example is far from conjectural. The type of coupling envisaged here is exactly that adopted by skilled marksmen in what Arutyunyan et al. (1968) call the "synergism of aiming."

Figure Captions

Figure 1. Finger positioning device showing programmable torque motors capable of producing a maximum torque of 81.6 oz-in about the metacarpophalangeal joint. Positioning errors were read directly from the pointer as it moved over the protractor graduated in degrees. Amount of torque to produce a perturbation could be varied as a percentage of maximum and the trigger point for the onset of perturbation was also controllable.

Figure 2. Acquisition trends indicate a plateau effect after 15 to 20 trials with a mean absolute error of about 2.3 degrees.

Figure 3. Actual displacement curves from a single subject in non-perturbed trials indicate three kinds of oscillatory behavior about the equilibrium point. Similar variability in the movement patterns of most subjects provides evidence that the stiffness parameters of the agonist and antagonist muscles were variable but that the ratio of activity between them was constant.

Figure 4. EMG recorded via surface electrodes placed over the extensor digitorum indicate an increase in the duration of extensor activity under wrist cuff conditions. As associated increase in activity of the flexor digitorum superficialis (not shown) indicates greater isometric influence. Movements were consequently slower (compare a and c, b and d), and without exception critically damped (c, d). The effect of perturbation in the wrist cuff condition (d) was less pronounced--a finding consistent with an increased overall stiffness in a mass-spring system perturbed by a constant load.

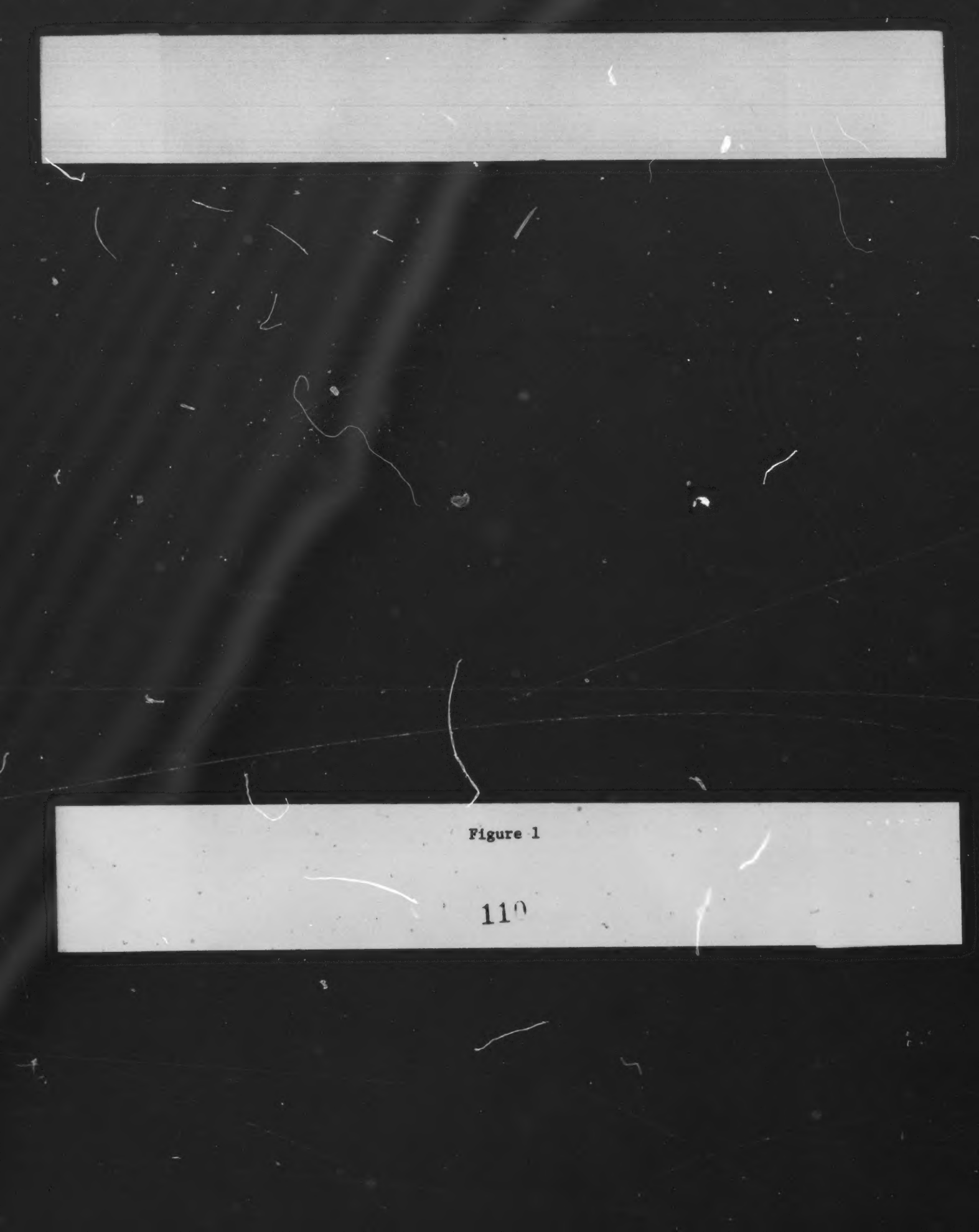


Figure 1

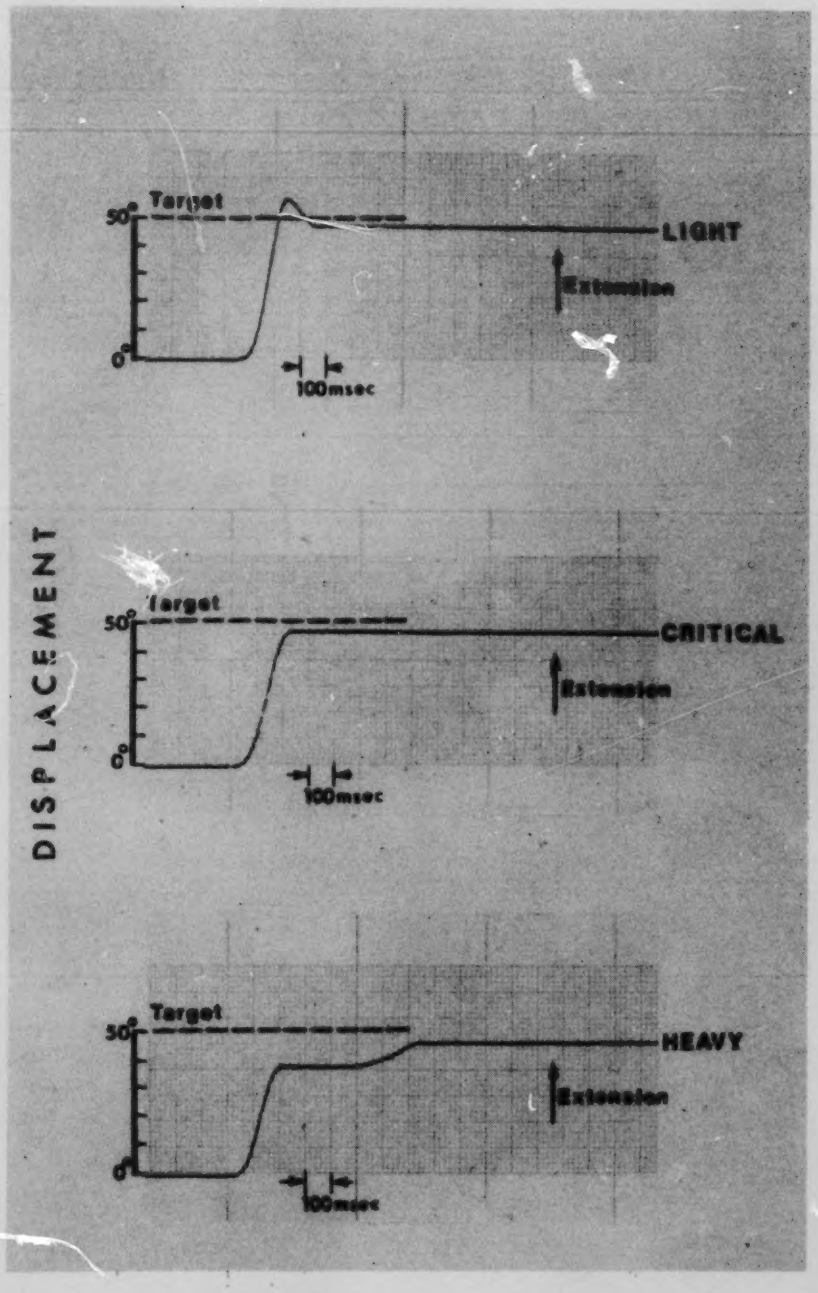

DISPLACEMENT
Target
50°
0°
LIGHT
Extension
100msec
Target
50°
0°
CRITICAL
Extension
100msec
Target
50°
0°
HEAVY
Extension
100msec

Figure 2

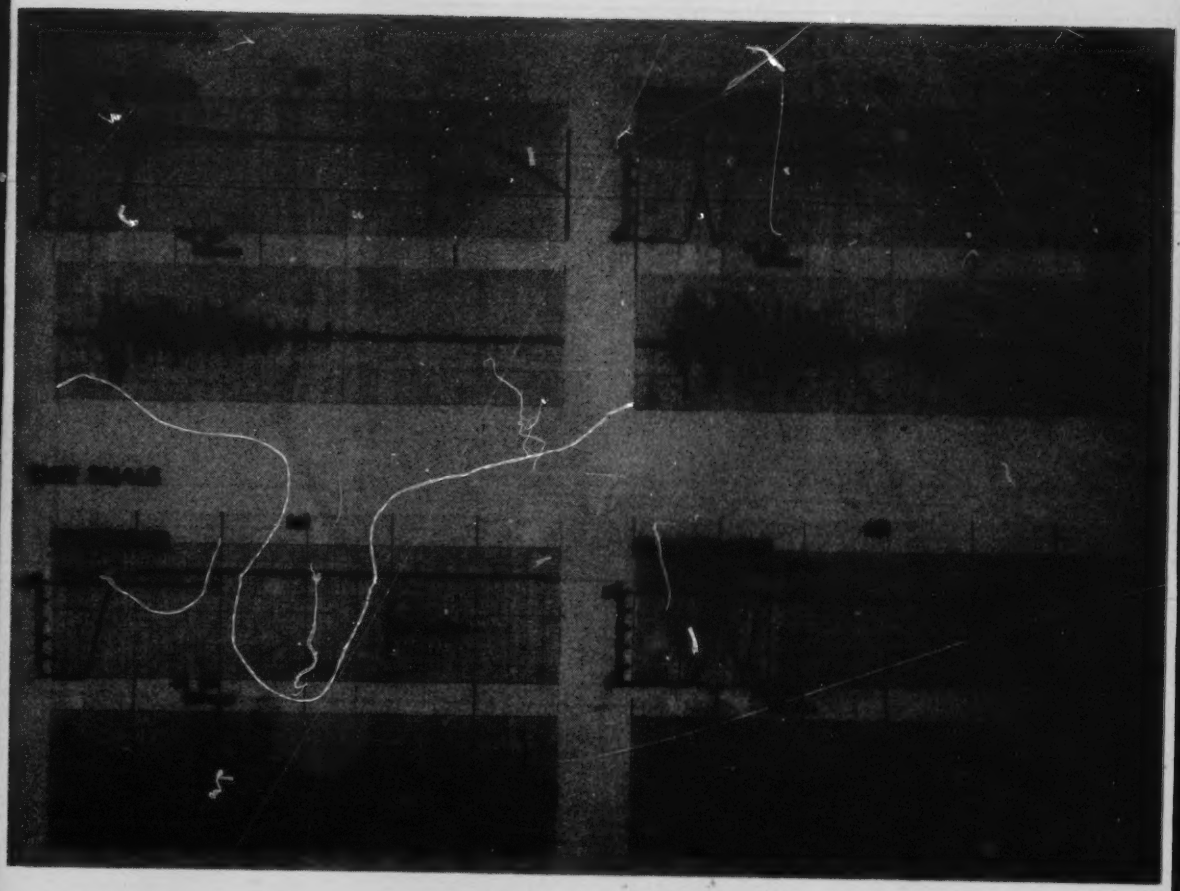


Figure 3

(This Figure accompanies the
immediately preceding paper.)

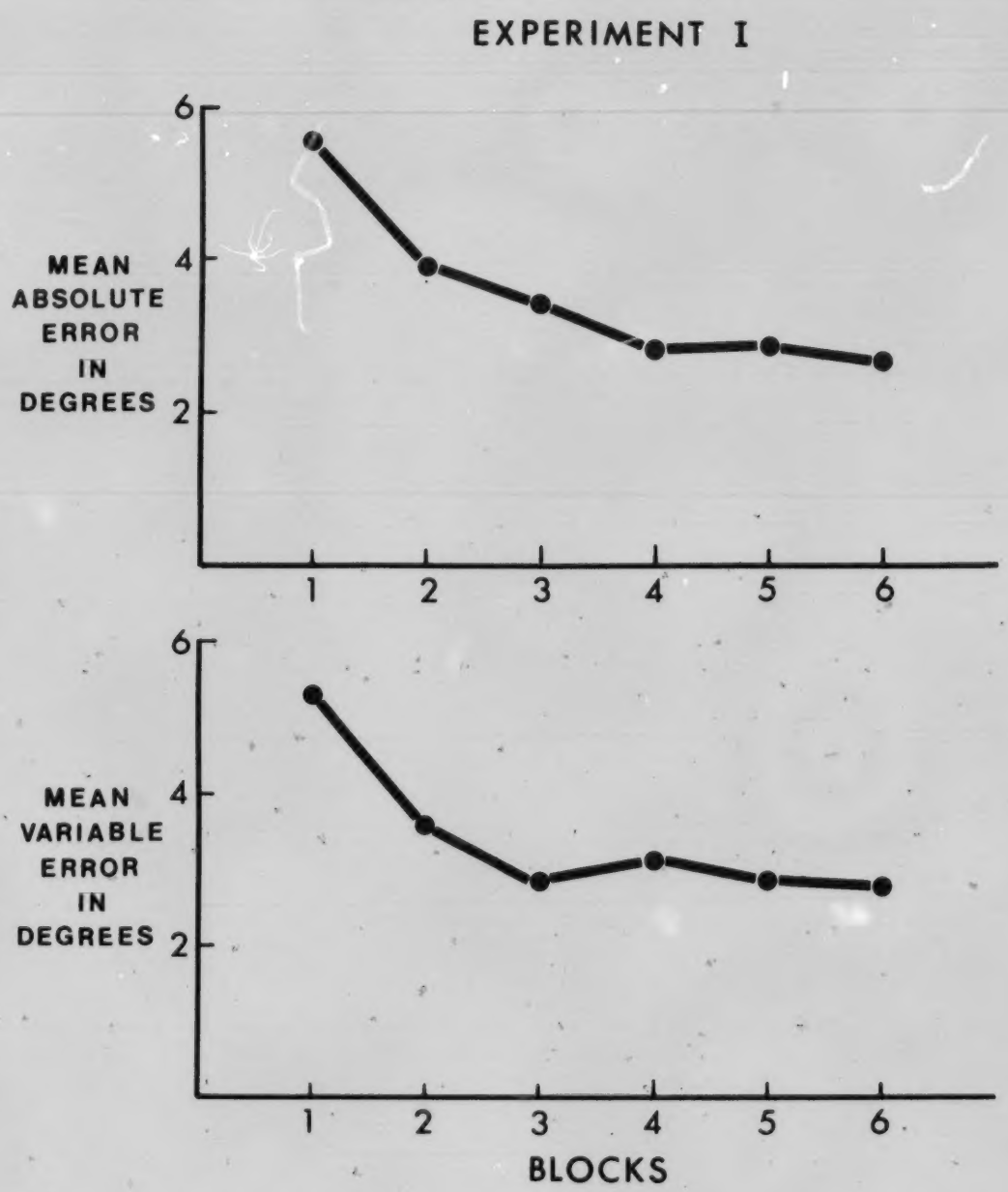


Figure 4

PROPERTIES OF SLOWLY ADAPTING JOINT RECEPTORS DO NOT READILY PREDICT PERCEPTION OF LIMB POSITION*

Wynne A. Lee+ and J. A. Scott Kelso++

Abstract. A passive limb position recognition task was used to examine two predictions based upon a neurophysiological model of slowly adapting (SA) joint receptors. First, we hypothesized that durations at target position coinciding with dynamic (3 and 6 sec) and static (15, 30 and 150 sec) phases of SA joint response should be associated with different accuracy levels favoring the longer durations. Second, because of the greater density of SA joint receptors sensitive to extreme positions, we predicted greater accuracy there than at intermediate limb positions. Results did not support either prediction about recognition accuracy. Errors for brief and long target durations were not statistically different. Increasing limb angles led to increased errors negating our second prediction. These data conflict with recent theorizing on the behavioral significance of joint receptors and indicate that knowledge of limb position is not readily predicted from joint receptor firing functions.

Recently a number of articles have argued that perception of limb position cannot be based entirely on input from slowly adapting (SA) joint receptors. Both behavioral (Goodwin, McCloskey, & Matthews, 1972; Roland, 1978) and neurophysiological (for review, Goodwin, 1976; Matthews, 1977) data have implicated cutaneous and muscle afferents as playing a role in perception of position. The authors of these studies attempt to manipulate the type of input available to subjects by selective or total occlusion of joint, skin, or muscle afferents; common techniques used have been reversible chemical (e.g., Roland, 1978), ischaemic (e.g., Kelso, 1977a), and thermal anesthesia of the hand and forearm (e.g., Paillard & Brouchon, 1974). Additionally, there are several studies of patients who have had joints (and, necessarily, joint receptors) surgically removed (Grigg, Finerman, & Riley, 1975; Kelso, Note 1). The consensus of these studies is that under conditions of selective anesthesia, perception of limb position is not entirely lost and that other sources of kinesthetic information may contribute to position sense.

---

*In Journal of Human Movement Studies, 1979, 5, 171-181.

+Also University of Texas at Austin

++Also University of Connecticut at Storrs

Acknowledgment: This research was supported by Grants AM-25814 and NS13617 from the National Institutes of Health.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

Usually, however, the conclusions of such studies are amended with the caution that the results obtained may not always be applicable to more normal conditions. Consequently, the belief that SA joint receptors may provide the normal basis for perception of limb position has been, and continues to be, fairly common (e.g., Adams, 1977; Marteniuk, 1976; Monster, Herman, & Altland, 1973; Russell, 1976). This belief is based on a 'classical' model of SA joint receptor function described in several reviews (e.g., Skoglund, 1973; Somjen, 1972) and the accompanying assumption that understanding of neurophysiological characteristics of sensory receptors would permit some prediction of perceptual-motor behavior under normal conditions.

This paper provides an initial attempt to test the assumption that knowledge of SA joint receptor characteristics enables prediction of limb position recognition accuracy under relatively normal conditions. First, the 'classical' model of SA joint receptor characteristics is presented, followed by significant modification of that model necessitated by recent neurophysiological research. Then, on the basis of the revised model of SA joint receptor characteristics, two specific predictions are made about recognition accuracy in a passive limb positioning task.

SA joint receptor characteristics. The 'classical' model has three features relevant to the idea that SA joint receptors are crucial to position sense:

(1) The full range of limb positions is supposedly represented equally throughout the population of SA receptors at a joint;
(2) The frequency responses of SA joint receptors are invariantly specific for a given angle;
(3) The angle-specific frequency responses are attained very rapidly (less than 2 seconds) and are steadily maintained as long as limb position is held.

More recent data challenge all three features of the classical model. In the first place, relatively few SA joint receptors fire at intermediate angles in cat and monkey (the common experimental animals); instead, most fire at full extension or flexion (Burgess & Clark, 1969; Clark, 1975; Grigg, 1975; Millar, 1975). Second, previous presentation of a joint angle influences subsequent receptor frequency responses to that angle. This hysteresis-like effect has been observed if a joint angle is maintained for as little as 30 seconds and may last as long as several minutes (McCall, Farias, Williams, & BeMent, 1974; Millar, 1975). Finally, SA joint receptors display two separable response phases to movement toward a subsequently maintained joint position (Mountcastle & Powell, 1959; Skoglund, 1973; McCall et al., 1974). The initial, dynamic phase is sensitive to velocity and acceleration as well as to displacement; this phase lasts approximately 12-15 seconds in unanesthetized preparations; in anesthetized preparations, abnormally rapid rates of adaptation are observed (Mountcastle, Poggio, & Werner, 1963), which may account for some reports of brief (1-2 sec) dynamic phase durations. The second, static (completely adapted) phase is position-sensitive only; it has been observed to last as long as ten minutes.

Model implications. The revised model of SA joint receptors raises some questions about results of previous studies that have been interpreted in

terms of theoretical notions of limb position accuracy, based on older neurophysiological data (for review, see Adams, 1977; Kelso, 1978). For example, if one assumes that stronger receptor responses are related to perceptual accuracy, then the greater responsiveness of SA joint receptors to extreme angles is not easily reconciled with findings that recognition of positions requiring greater movement extent is less accurate than recognition of positions at intermediate or short extents. Rather, one would expect greater accuracy at either extreme joint position than at intermediate locations. It may be that positioning studies have tended to use limb positions defined so that shorter (more accurate) movements were at extreme angles, and longer (less accurate) movements were at intermediate angles. The first hypothesis that the present study is designed to test is that smaller recognition errors should obtain for both extreme flexion and extension positions than for intermediate positions, since the relative density of SA joint receptors is higher at the ends of the motion range.

The existence of a dynamic as well as static response phase for SA joint receptors poses additional interesting problems for studies of positioning accuracy. With a single exception, all previous positioning studies have used durations on target of 0 to 12 seconds, a period well within the dynamic rather than static response range. One might argue that position accuracy at such brief target durations cannot logically be attributed to the capability of SA joint receptors for maintaining a steady firing frequency as long as position is held. Indeed, two studies (Monster et al., 1973; Paillard & Brouchon, 1968) that used durations within the dynamic response interval reported increasingly negative constant error with increased target durations. Wallace and Stelmach (1975) and Stelmach and McCracken (1978) found changes in absolute error between 0 and 2 or 5 second durations but no difference between 2 and 5 seconds. Del Rey and Lichter (1971) observed no changes in error scores with target durations of 2, 5, and 10 seconds. These results suggest somewhat variable position recognition for target durations selected from the 'dynamic' response-phase range of SA joint receptors. In contrast, the one study that used longer target durations (15, 60, 180 sec, which would coincide with the static SA joint receptor response phase) reported no difference in recognition errors over the three durations, as might be predicted from knowledge of the steady firing of receptors during that period (Horsch et al., 1975).

However, no experimental comparison of position accuracy between target durations within the dynamic and static response phases has yet been attempted. If the steadily maintained frequency response of SA joint receptors provides optimal information about position, then one might predict greater accuracy for target durations within the static phase. Receptor input arising from briefer target durations might be confounded with velocity and acceleration information. Thus, the second purpose of the present study is comparison of position accuracy for brief (3, 6 sec) and long (15, 30, 150 sec) target durations. If the dynamic/static response phase distinction of SA joint receptors has predictive value for perception of position, then accuracy differences between the two classes of target durations should obtain. Moreover, there should be no difference in accuracy among the three longer durations; that is, the findings of Horsch, Clark, and Burgess (1975) should be replicated. The durations of 15, 30, and 150 sec were selected, in part, to test the latter notion. A period of 30 sec was chosen to provide a time at

which reflex and mechanical-elastic responses of the muscles to the targeting movement should be minimal (Hayes & Hatze, in press). The 150 sec target duration was used as a time at which conscious awareness of position due to cutaneous inputs should be unavailable to subjects, since appreciation of constant skin indentation fades after one to two minutes (Horsch et al., Experiment 2, 1975). Thus, accuracy differences among these three durations might suggest that either muscle or cutaneous afferents might be contributing to position sense.

There is an additional difficulty in comparing errors in position accuracy when target durations vary. As Wallace and Stelmach (1975) and Horsch et al. (1975) pointed out, central as well as peripheral explanations can be used to account for similarities or differences in positional accuracy associated with varied target durations. Amount of time to attend to target position, or to rehearse recognition of the position will increase with increased target duration. Without some control for potential rehearsal or attention effects, there is no empirical reason to choose either the peripheral (receptor) or central hypothesis concerning accuracy changes with varied target duration. In the present experiment, an attempt was made to control for the amount of time that subjects could pay full attention to a target position. A secondary, masking task was performed with the left arm during passive positioning of the subject's right arm by the experimenter. One group of subjects (non-masked) performed the secondary (mirror tracing) task only during movements toward and away from the target position. The other group of subjects performed the tracing task during movements and also throughout the target duration, except for the final three seconds when they were cued to pay attention to the position of the right arm. Thus, subjects in the masked group were restricted to 3 sec during which they could give undivided attention to the target position; subjects in the non-masked group were free to attend fully to the target position for the entire target duration. There is no reason to suspect that the mirror tracing task was wholly adequate in blocking subjects' attention to arm movement or position. However, emphasis on accuracy in the tracing task (which was performed with the non-preferred hand) was strong, the task was continuous, and it required integration of visual-proprioceptive input. The combination of these three characteristics made mirror tracing a reasonable choice as a secondary task that would require attention and would interfere structurally with the information required for the positioning task.

The primary task consisted of an angular, horizontal positioning movement involving the right elbow joint with vision of the positioning device excluded. Individuals were asked to indicate verbally the position of the right elbow, which had previously been presented as a target. The sensitivity of this technique has been demonstrated in previous work (Kelso, 1977b). A passive mode of presenting both target and recognition positions was adopted. In active movements, central and peripheral modulation of SA joint receptor inputs would be more likely than in passive movements. Therefore, the passive task should maximize the contribution of SA joint receptor input to the perception of limb position.

Method

Subjects

Thirty right-handed, adult volunteer subjects were assigned in alternating sequence to one of the two experimental groups. There were 8 males and 7 females in each group.

Apparatus

The angular positioning device, placed in the horizontal plane, had angles marked off along the periphery in one degree increments (0 degrees at base left, 180 degrees at base right). A padded lever arm pivoted freely about a shaft centered at the base of the device. The apparatus was hidden from the subject's view by a curtain, with an opening in the curtain large enough for the subject's right arm. A sling was suspended above the base of the lever; the sling stabilized the position of the subject's upper arm, allowing the arm to pivot at the elbow. Two Velcro straps maintained forearm position on the lever. A height-adjustable stool was used for seating. The subjects were situated such that the right arm was fully extended (180 degrees) on the lever arm when the lever arm pointed to 145 degrees (at the periphery), with the arm abducted 90 degrees from the body without shoulder protraction or retraction. The range of motion used in the experiment was from 45 degrees to 135 degrees (10 degrees to 135 degrees respectively on the device). The relatively large 'extreme flexion' angle of 45 degrees was selected because in pilot work it was found that several persons were uncomfortable with greater degrees of flexion. Errors were recorded to the nearest .5 degrees.

A mirror-tracing task (star-shaped) was used for the secondary masking task. The tracing apparatus was placed at a right angle to the left edge of the positioning device. The subjects performed the star-tracing task left-handed, with the eyes and head turned toward the tracing apparatus at about 45 degrees to left of body center. They were instructed to maintain this head position throughout the experiment. Performance was measured as the number of times off target (measured by an electronic counter) per star traced for each block of trials.

A stopwatch was used to time all durations within the experiment. All experimenter-produced movements were timed, to permit maintenance of a constant average velocity of 10 degrees per second.

Procedure

The subjects were told that the experiment tested their ability to perform as well as possible on two different tasks. Instructions emphasized that optimal performance required full attention on one task at a time and that dividing attention would lead to poorer overall performance. A bonus of $5 was used to motivate subjects to maximize performance in both tasks; the bonus was awarded to the subject with the best combined score from the star-tracing and the position recognition tasks. Following verbal presentation of the instructions for the two tasks, all subjects received practice trials until they demonstrated an ability to keep the right arm relaxed and to follow

the experimenter's commands correctly; no subject required more than four practice trials.

The subjects began each trial with the arm at 45 degrees (flexed position) and with their eyes shut. On the command "ready," subjects opened their eyes; one sec later they were told to begin tracing. After 5 sec, the experimenter began moving the subject's right arm to the target location, while the subject continued the tracing task.

Non-masked group. When the target angle was achieved, subjects in the non-masked group were told, "Stop, attention arm"; at this point, timing of the target duration began. The command meant that the subject should stop tracing, close his/her eyes, and quickly shift attention to the position of the right arm. The subject was instructed to sit quietly and to continue paying attention to the arm position until hearing the command, "Trace again." This command was given 1 sec before completion of the target duration; at completion of the duration, when the subject had resumed tracing, the experimenter moved the arm back to the start position at 45 degrees. After 15 sec from the beginning of the return movement, the experimenter said, "Stop, recognize," at which time the subject again closed his eyes, stopped tracing, and attended to his right arm, which the experimenter immediately began to move from the start position toward the target location. When the subject felt his arm was in the same position that he attended to during the criterion presentation, he indicated recognition by saying, "Stop." The subjects were instructed to tell the experimenter to move the arm forward or back, if the experimenter had not stopped in the position that the subject had indicated. After the experimenter recorded the angle that the subject indicated, the subject's arm was moved back to the start position. The intertrial interval was 15 seconds.

Masked group. For subjects in the masked group, procedures were essentially the same except that they continued tracing during all but the last 3 sec of the target duration. Three sec before the end of the target duration, they were told to attend to the arm; two sec later they were told to "Trace again." At the completion of the target duration, the experimenter moved the arm back to the start position. The remainder of the trial was exactly the same for the masked group as for the non-masked condition.

Design

A mixed 2 x 3 x 5 factorial design was used, with repeated measures on the second two factors. Secondary task activity (masked, non-masked) was the between-groups variable. Response sector (flexion, extension, and intermediate) and target duration (3, 6, 15, 30, 150 sec) were the within-subject variables. Absolute, constant, and variable error (AE, CE, and VE, respectively) were the dependent measures. The subjects in each group (n = 15) received 60 trials on the passive, angular recognition task; four trials were given in each response sector for all five target durations. Four blocks of 15 trials were presented to the subjects in two sessions that were at least 24 hours apart. In every block, all five target durations were presented three times in randomized order; each duration was matched with one angle for each response sector. There were 12 criterion angles, 4 per response sector (flexion = 55, 58, 61, and 64 degrees; intermediate = 106.5, 109.5, 112.5, and

115.5 degrees; extension = 161, 164, 167, and 170 degrees). Consequently, in each block of trials three angles were repeated once. Two conditions were imposed on these repetitions to minimize possible hysteresis-like effects on joint receptors adapted responses: (1) At least four trials intervened between repeated angles; (2) In no block was a repeated angle presented first at the 150 sec duration.

Results

Recognition Errors

Although errors for the non-masked group were generally lower than those for the masked group, the difference reached significance only for AE, $F(1,28) = 5.62$, $p < .05$. Mean absolute, constant, and variable errors for groups are presented as a function of target duration in Table 1.

Table 1

Mean Absolute, Constant, and Variable Error (in degrees) for Masked and Non-Masked Groups under Different Target Durations

	Duration (seconds)				
Group	3	6	15	30	150
Masked					
AE	8.34 (3.95)*	8.82 (3.01)	9.06 (3.52)	9.37 (3.64)	9.68 (4.34)
CE	-5.40 (5.20)	-5.03 (4.75)	-5.44 (4.99)	-6.05 (5.06)	-6.02 (6.23)
VE	4.78 (2.07)	5.19 (2.80)	5.17 (2.46)	5.45 (2.83)	4.33 (2.06)
Non-Masked					
AE	7.46 (3.06)	8.48 (3.20)	7.59 (3.45)	7.32 (3.55)	6.20 (3.49)
CE	-4.37 (4.46)	-5.33 (4.14)	-4.26 (4.34)	-4.51 (4.46)	-2.28 (4.87)
VE	4.89 (1.94)	5.29 (2.69)	4.72 (1.95)	4.27 (2.01)	4.49 (2.21)

AE = Absolute Error, CE = Constant Error, VE = Variable Error
*Standard deviations of error scores are in parentheses

The main effect for duration-at-target did not reach the .05 level of significance for AE, CE, or VE, F (4,112) = .87, 1.30, and 1.35, respectively. Thus the hypothesis that the times associated with the dynamic and steady adapted response phases of receptors would be associated with different error levels was not supported.

However, the duration-by-groups interaction was significant for AE, F (4,112) = 4.32, $p < .01$, and for CE, F (4,112) = 3.54, $p < .01$, though not for VE, F (4,112) = 1.24, $p > .05$. Newman-Keuls comparisons between the groups for each duration revealed a significant difference between groups only at 150 sec for AE, q (10,50) = 3.48, $p < .01$. For CE, none of the individual group by duration comparisons reached the .05 level of significance. As may be seen in Table 1, AE and CE showed essentially the same pattern. Errors were higher generally for the masked group, and differences between groups tended to increase for the longer durations.

Table 2

Mean Absolute, Constant, and Variable Errors (in degrees) within Three Target Sectors for Masked and Non-Masked Groups

	Target Sector		
Group	Flexion	Intermediate	Extension
Masked			
AE	4.85	10.16	12.04
	(2.26)*	(4.10)	(4.71)
CE	2.60	-8.24	-11.12
	(3.72)	(6.06)	(5.95)
VE	3.90	5.40	5.66
	(1.90)	(2.96)	(2.49)
Non-masked			
AE	4.21	7.65	10.39
	(2.44)	(3.59)	(4.07)
CE	1.99	-4.83	-9.61
	(3.53)	(5.05)	(4.79)
VE	3.36	5.87	4.96
	(1.68)	(2.66)	(2.13)

*Standard deviations of error scores are in parentheses

Mean absolute, constant and variable errors are presented as a function of target sector in Table 2. The sector main effect was highly significant, F (2,56) = 56.09, $p < .01$, F (2,56) = 158.04, $p < .01$, and F (2,56) = 20.87, $p <$

.01 for AE, CE and VE, respectively. The range effect was clearly evident, with overshooting characterizing the flexion sector ($\underline{M}$ = 2.30 degrees), moderate undershooting for the intermediate sector ($\underline{M}$ = -5.68 degrees), and highest negative error for the extension sector ($\underline{M}$ = -10.37 degrees). AE increased from flexion to intermediate to extension sectors ($\underline{M}$ = 4.53, 8.91, and 11.21 degrees, respectively). For VE, the flexion sector showed the smallest average error ($\underline{M}$ = 3.63 degrees), the intermediate sector showed the largest average error ($\underline{M}$ = 5.63 degrees), and the extension sector error was slightly lower than that for the intermediate sector ($\underline{M}$ = 5.31 degrees). Thus, VE was the only error score showing better performance in both extreme sectors. However, since overall accuracy (AE) indicated that errors increased as target angle (and movement extent to the target) increased, the data on VE must be deemed insufficient as evidence for greater accuracy at extreme limb positions.

The sector-by-group interaction was significant only for CE, $\underline{F}$ (2,56) = 3.80, $\underline{p} < .03$. For each sector, errors were closer to 0 degrees for the non-masked than the masked group. The duration-by-sector interaction failed to reach the .05 level of significance for AE, CE, or VE, $\underline{F}$s (8,224) < 1. The three-way duration-by-sector-by-groups interaction was significant for AE, $\underline{F}$ (8,224) = 2.61, $\underline{p} < .01$, and for CE, $\underline{F}$ (8,224) = 2.41, $\underline{p} < .05$; neither of these interactions was readily interpretable.

Star-Tracing Errors

For the secondary task, measures of error rate (number of errors divided by the number of stars completed per block) were compared on the first and fourth blocks of trials, using $\underline{t}$-tests for correlated means. These comparisons were performed to ensure that subjects had improved across blocks of trials, on the assumption that improvement would indicate that subjects were indeed continuing to pay attention to the tracing task. The differences proved significant for both the masked and non-masked groups, $\underline{t}$ (14) = 5.25, $\underline{p} < .01$, and $\underline{t}$ (14) = 2.27, $\underline{p} < .025$, respectively. For the masked group, error rate scores showed a consistent decrease with blocks ($\underline{M}_1$ = 15.19, $\underline{M}_2$ = 7.5, $\underline{M}_3$ = 6.06, $\underline{M}_4$ = 4.57 errors/star). For the non-masked group, there was a similar trend, except that block 3 (the first block of the second session) had slightly higher errors than block 2 ($\underline{M}_1$ = 11.85, $\underline{M}_2$ = 7.07, $\underline{M}_3$ = 7.70, $\underline{M}_4$ = 6.29 errors/star).

DISCUSSION

The data of this study did not support the hypothesis that accuracy in recognition of a passively presented limb position might be predicted from known neurophysiological characteristics of slowly adapting joint receptors. The first prediction, that errors should be smaller for the flexion and extension sectors than the intermediate sector, was not confirmed. It is beyond the scope of these data to speculate on the nature (either central or peripheral) of the processes that might cause increased error with increased distance moved to a target position (but see Weiss, 1954; Wilberg & Girouard, 1975). What is clear is that a chief property of SA joint receptors, namely their differential response to extreme versus intermediate joint angles, did not facilitate accurate prediction of recognition errors in limb positioning.

The second hypothesis, that times corresponding to dynamic and static phases of SA joint receptor responses would be associated with different error levels, was also largely unsupported by the data of this experiment. There was no evidence that the longer target durations (presumed to coincide with the static, steady adapted response phase) allowed subjects to more accurately recognize the target position. The dynamic/static response phase dichotomy, evident in neurophysiological studies, was apparently without a behavioral counterpart under the conditions of low velocity, passive movements used in this study. However, as Burgess and Perl (1973) pointed out, for displacements of low acceleration or velocity, the dynamic response of mechanoreceptors provides a relatively undistorted position signal. That is to say, the movement velocities used in this study may have been too small to elicit a dynamic phase response large enough to significantly disrupt positional information in the receptor response. A more appropriate test of the behavioral significance of the dynamic and static response-phases would be to use movements of high acceleration or velocity. However, within the restricted velocity rate used in the present study, knowledge of dynamic and static response-phases was not predictive of perceptual accuracy for limb position.

These results may be compared with those of other studies that have also used low or moderate velocity movements in positioning tasks. For the durations associated with the dynamic response phase, the data from this study were consistent with those of Del Rey and Lichter (1971), Wallace and Stelmach (1975), and Stelmach and McCracken (1978); that is, no difference between recognition errors for 3 and 6 seconds was observed. These results contrast with those of Paillard and Brouchon (1968) and Monster et al. (1973) who reported a consistent tendency for constant error to become more negative when time at target was increased from zero to 12 seconds. Those authors suggested that the constant error effect might be related to the decay of the velocity-sensitive dynamic response of slowly adapting joint receptors. That hypothesis seems somewhat questionable since the studies noted above found no constant error effect. More likely the discrepancy is related to a difference in experimental procedures. Both Paillard and Brouchon (1968) and Monster et al. (1973) used a task in which the vertical position of one limb was matched by the opposite limb. The current study and those by Del Rey and Lichter (1971), Wallace and Stelmach (1975), and Stelmach et al. (1975) all used a horizontal task in which the criterion and reproduction movements were performed with the same limb in the horizontal plane. Either the difference in the plane of movement or in the use of the same or opposite limb to match the criterion position might account for the discrepancy in results.

The data showing statistically equivalent errors at longer 'static phase' target durations are in agreement with those reported by Horsch et al. (1975). The argument that rehearsal time might have been a critical factor in equating the three durations was not substantiated, for errors did not change significantly with durations for either the masked or non-masked group. Moreover, one cannot attribute the superiority of the non-masked group simply to the longer time available to attend to limb position. Were attention time crucial per se, two results should have been observed. First, the superiority of the non-masked over the masked group should have increased consistently with longer durations. This prediction received only marginal support in the one significant difference between groups at 150 sec (AE). Second, within the

non-masked group, errors should have decreased significantly with increased time on target; some decline was noted, but it was not significant. An alternative explanation for the differences between the masked and non-masked groups may be that the secondary tracing task interfered with the recognition task, possibly by fatiguing subjects, especially at the long durations. An overall fatigue factor does not seem supported: Error differences between the two groups at 3 and 6 secs were very small.

As noted in the introduction, the present study represented an initial attempt to test directly the notion that perception of position can be predicted from knowledge of slowly adapting joint receptor firing functions. It would be useful to perform additional, more stringent tests of the relationship between receptor properties and perception--by using movements of higher velocity for example--or by monitoring muscle activity to ensure passivity, or by recording nerve impulse volleys from joint receptors during a positioning task. However, within the limitations of this study, the main conclusion was clear: There was not a straightforward relationship between accuracy of limb position perception and the two specified neurophysiological properties of SA joint receptors, even under 'passive' conditions.

This conclusion may not be surprising, since available neurophysiological data are based on experiments with subhuman species. Firing functions for human SA joint receptors have yet to be studied. It also seems likely that higher processes could significantly modify receptor input or the interpretation of receptor input according to the needs of the information processing system. For example, accuracy of limb position perception can be improved with knowledge of results (for review, see Newell, 1977); presumably, receptor input remains basically unchanged but use of that information is improved with learning.

Given the problems with extrapolating from neurophysiological properties of receptors in subhuman species to human perceptual behavior, it is intriguing that investigators frequently attribute findings to the functioning of specific receptors. A case in point is the often replicated finding that memory for limb position or 'location' is superior to memory for movement amplitude or 'distance' (for review see Stelmach & Kelso, 1977). The fact that there are SA joint receptors that can signal location but not distance is often cited as a reason for the difference in memory for these two qualities. The present data, as well as other evidence (Kelso, 1978 for review), suggest that it is time to reconsider such an interpretation. Similarly, accounts of other active movement behaviors, such as motor timing based solely on the properties of SA joint receptors (Adams, 1977), seem unrealistic. At least for accuracy in discrete positioning movements, the need for peripheral input is greatly reduced under active conditions (Grigg et al., 1973; Kelso, 1977a, 1978; Polit & Bizzi, 1979). Hence, if even passive position sense cannot be readily predicted from knowledge of SA joint receptor properties, then it seems unlikely that active movement behavior should be attributable entirely to these receptors' functional properties.

Despite the problems inherent in trying to make behavioral predictions from a neurophysiological model of receptor function, such attempts may be justified in at least two ways. First, data from studies making deliberate predictions can reveal how proposed neurophysiological explanations of

behavior may be overly simplistic. Second, such studies may lead to potentially more accurate models interfacing behavioral and neurophysiological levels of description, because they can indicate which behaviors can be explained with reference to receptor level events and which cannot, thereby suggesting the need for more complex explanatory models.

REFERENCE NOTE

1. Kelso, J. A. S. Finger localization following total joint arthroplasty. Paper given at Psychonomic Society Meetings, San Antonio, Texas, 1978.

REFERENCES

Adams, J. A. Feedback theory of how joint receptors regulate the timing and positioning of a limb. Psychological Review, 1977, 84, 504-523.

Burgess, P. R., & Clark, F. J. Characteristics of knee joint receptors in the cat. Journal of Physiology, 1969, 203, 301-317.

Burgess, P. R., & Perl, E. R. Cutaneous mechanoreceptors and nociceptors. In A. Iggo (Ed.), Handbook of sensory physiology, Vol. 11: Somatosensory system. Berlin: Springer-Verlag, 1973, 29-78.

Clark, F. J. Information signaled by sensory fibers in medial articular nerve. Journal of Neurophysiology, 1975, 38, 1464-1472.

Del Rey, P., & Lichter, J. Accuracy in horizontal arm positioning. Research Quarterly, 1971, 42, 150-155.

Goodwin, G. M. The sense of limb position and movement. In J. Keogh & R. S. Hutton (Eds.), Exercise and sport sciences reviews (Vol. 4). Santa Barbara, Calif.: Journal Publishing Affiliates, 1976.

Goodwin, G. M., McCloskey, D. I., & Matthews, P. B. C. The contribution of muscle afferents to kinaesthesis as shown by vibration induced illusions of movement and by the effects of paralyzing joint efferents. Brain, 1972, 95, 705-748.

Grigg, P. Mechanical factors influencing response of joint afferent neurons from cat knee. Journal of Neurophysiology, 1975, 38, 1473-1484.

Grigg, P., Finerman, G. A., & Riley, L. H. Joint-position sense after total hip replacement. Journal of Bone and Joint Surgery, 1973, 55A, 1016-1025.

Hayes, K. C., & Hatze, H. Passive visco-elastic properties of the structures spanning the human elbow joint. European Journal of Applied Physiology, in press.

Horsch, K. W., Clark, F. J., & Burgess, P. R. Awareness of knee joint angle under static conditions. Journal of Neurophysiology, 1975, 38, 1436-1447.

Kelso, J. A. S. Motor control mechanisms in human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 529-543. (a)

Kelso, J. A. S. Planning and efferent components in the coding of movement. Journal of Motor Behavior, 1977, 9, 33-47. (b)

Kelso, J. A. S. Joint receptors do not provide a satisfactory basis for motor timing and positioning. Psychological Review, 1978, 85, 474-481.

Marteniuk, R. G. Information processing in motor skills. New York: Holt, Rinehart, Winston, 1976.

Matthews, P. B. C. Muscle afferents and kinaesthesia. British Medical Bulletin, 1977, 33, 137-142.

McCall, W. D., Farias, M. C., Williams, W. J., & BeMent, S. L. Static and dynamic responses of slowly adapting joint receptors. Brain Research, 1974, 70, 221-243.

Millar, J. Flexion-extension sensitivity of elbow joint afferents in cat. Experimental Brain Research, 1975, 24, 209-214.

Monster, A. W., Herman, R., & Altland, N. R. Effect of the peripheral and central 'sensory' component in the calibration of position. In J. E. Desmedt (Ed.), New developments in electromyography and clinical neurophysiology (Vol. 3). Basel: S. Karger, 1973, 383-403.

Mountcastle, V. B., Poggio, G. F., & Werner, G. The relation of thalamic cell response to peripheral stimuli varied over an intensive continuum. Journal of Neurophysiology, 1963, 26, 807-834.

Mountcastle, V. B., & Powell, T. P. S. Central nervous mechanisms subserving position sense and kinesthesis. Bulletin of the Johns Hopkins Hospital, 1959, 105, 173-200.

Newell, K. Knowledge of results and motor learning. In J. Keogh & R. S. Hutton (Eds.), Exercise and sports science review (Vol. 4). Santa Barbara, Calif.: Journal Publishing Affiliates, 1977, 195-228.

Paillard, J., & Brouchon, M. Active and passive movements in the calibration of position sense. In S. J. Freedman (Ed.), The neuropsychology of spatially oriented behavior. Homewood, Ill.: Dorsey Press, 1968, 37-56.

Paillard, J., & Brouchon, M. A proprioceptive contribution to the spatial encoding of position cues for ballistic movements. Brain Research, 1974, 71, 273-284.

Polit, A., & Bizzi, E. Characteristics of the motor programs underlying arm movements in monkeys. Journal of Neurophysiology, 1979, 42, 183-194.

Roland, P. E. Sensory feedback to the cerebral cortex during voluntary movement in man. The Behavioral and Brain Sciences, 1978, 1, 129-147.

Russell, D. G. Spatial location cues and movement production. In G. E. Stelmach (Ed.), Motor control: Issues and trends. New York: Academic Press, 1976, 67-86.

Skoglund, S. Joint receptors and kinesthesis. In A. Iggo (Ed.), Handbook of sensory physiology, Vol. 11: Somatosensory system. Berlin: Springer-Verlag, 1973, 111-136.

Somjen, G. Sensory coding in the mammalian nervous system. New York: Appleton-Century Crofts, 1972.

Stelmach, G. E., & Kelso, J. A. S. Memory processes in motor control. In S. Dornic (Ed.), Attention and performance VI. Hillsdale, N.J.: Erlbaum, 1977.

Stelmach, G. E., Kelso, J. A. S., & Wallace, S. A. Preselection in short-term motor memory. Journal of Experimental Psychology: Human Learning and Memory, 1975, 6, 745-755.

Stelmach, G. E., & McCracken, H. Storage codes for movement information. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Erlbaum, 1978.

Wallace, S. A., & Stelmach, G. E. Proprioceptive encoding in preselected and constrained movements. Mouvement, 1975, 7, 147-152.

Weiss, R. B. The role of proprioceptive feedback in positioning responses. Journal of Experimental Psychology, 1954, 47, 215-224.

Wilberg, R. B., & Girouard, Y. On the locus of the range effect in a short term motor memory paradigm. Mouvement, 1975, 7, 153-162.

PERCEIVING PHONETIC SEGMENTS*

Michael Studdert-Kennedy+

Why do we study speech perception? Is the speech signal merely a complex acoustic structure in which we are interested because we happen to use it for communication? Certainly, speech belongs to a natural class of acoustic patterns about which we know very little, namely, patterns structured over time by mechanical events--such as a footstep on gravel, a hand crumpling paper, a glass bottle bouncing or breaking (Warren, personal communication). If we knew more about the temporally and spectrally distributed acoustic properties of such dynamic events, and how we recognize them, would we then understand how we perceive speech? In other words, is speech merely a distinctive set of sounds, shaped, as no other sounds are, by movements of human articulators modulating a characteristic vocal source? Perhaps.

But there are reasons to believe that speech may also be distinctive in some deeper sense than this, and that it may engage distinctive perceptual mechanisms. First, the speech signal is the primary carrier of a language: It is rich in information not only about the abstract phonological structure of an utterance, but also about its syntactic and semantic structure. To perceive speech is to apprehend this structure, an act that can only be accomplished by a listener who knows a language. Just what the listener knows is, of course, very much the question. But he must, at least, know the abstract linguistic units that compose the message--whether words, morphemes or phonemes--and he must know how such units are realized in the signal. No other natural sound, so far as we know, conveys so complex an abstract message.

A second and simpler reason for suspecting that the speech signal may be uniquely related to the human perceptual apparatus is that each speaker of a particular language can readily reproduce sounds uttered by another. Moreover, the speaker can repeat the sounds so rapidly that we may reasonably hypothesize a privileged relation between elements of perception and elements of production. That the repetition may be far from an acoustically exact imitation, and yet be perceived as a repetition, only strengthens the

---

*This chapter is a revised version of a paper given at the International Symposium on the Cognitive Representation of Speech, held in Edinburgh, July 29th through August 2nd 1979. It will appear in T. F. Myers, J. Laver, and J. Anderson (Eds.), The cognitive representation of speech, Amsterdam: North-Holland.

+Also at Queens College and the Graduate Center, City University of New York.

Acknowledgment: I thank Alvin Liberman for shrewd comments. Preparation of this chapter was supported in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

hypothesis by emphasizing that more than mere acoustic identity is involved.

We should not lightly dismiss the ability to repeat or imitate. Imitation is, in fact, a remarkable skill that requires an animal to parse the behavior of another into components, and then activate its own corresponding motor controls to reproduce the behavior. A general capacity to imitate has evolved only in social animals--for the most part, in certain primates: Its adaptive advantages are obvious in, for example, the well-known sweet potato washing of Japanese macaques (Wilson, 1975, pp. 170-171) or the nest-building of young chimpanzees (van Lawick-Goodall, 1971). Vocal imitation has evolved only in certain species of social marine mammals and of birds, and in humans: Its adaptive advantages can only lie in communication. Moreover, transformation from sensory input to corresponding motor control is less direct for an acoustic signal, shaped by movements within the internal space of a vocal tract, than for an optical signal, shaped by movements in the external space common to observer and observed. Perhaps we should not be surprised that the human infant begins its imitative attempts by tracking the visually available movements of its caretaker's mouth in "prespeech" oral play (Trevarthen, Hubley, & Sheeran, 1975). In any event, the human capacity to imitate the speech of another implies a specialized sensorimotor device (perhaps analogous to that being discovered in the canary [Nottebohm, 1977]) and suggests that phonetic structure may be represented in the brain in a form sufficiently abstract for ready interchange between listening and speaking. Whether this representation is necessarily elicited during normal speech perception is, of course, an open question.

From this introduction, I wish to turn to the question of how listeners parse an utterance into its component linguistic segments. Whether the segments are words, morphs, syllables or phones need not concern us, for the moment, since once we have accepted the onus of describing the relation between linguistic segments and the acoustic signal, the problem is essentially the same for all. However, I shall concentrate on the phone, and its constituent cues or features, for several reasons, more fully discussed elsewhere (Liberman & Studdert-Kennedy, 1978). First, I assume the abstract elements of phoneme or feature that underlie the phonetic string to be psychologically real, structural units of lexical and grammatical morphemes, essential to the dual structure that makes linguistic communication possible. Second, a solution to the segmentation problem at the phonetic level should lead toward a solution at higher levels. In fact, we already have some evidence that English word juncture is signaled by allophonic variations in the immediate vicinity of the juncture, that these variations occur at onset rather than offset of words (Nakatani & Dukes, 1977) and that acoustic-phonetic structure at word onset is particularly important for lexical access in fluent speech (Marslen-Wilson & Welsh, 1978). Finally, the speed and efficiency of speech as a medium of communication rests, in large part, on coarticulation among phonetic segments, and this coarticulation is the primary source of the segmentation problem (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

The trouble began with the spectrogram and with the recognition that, as the cliché has it, phones are not strung together likes beads on a string, but rather, as Hockett (1958) wrote, like a line of eggs passed between rollers. Historically, there have been two broad responses to the paradox. The first

has been to accept the segments of linguistic analysis--whether the features of distinctive feature theory or the static targets of traditional articulatory phonetics--as psychologically real, and then either to attempt an acoustic formulation of the signal isomorphic with those segments, or to posit some specialized mechanisms for linking signal and message. The second response has been frankly pragmatic, proposing to finesse the linguist's units as little more than analytically useful fictions (perhaps encouraged by the historical "accident" of the alphabet) and to go for supposedly simpler groupings such as phoneme dyads, syllables, or even words.

The first line of response has been the more varied and complex of the two, and will be the focus of most of what I have to say. An important contributor here was Fant who early recognized the disparity between signal and message, remarking that a single segment of sound may convey information concerning several segments of the message, and a single segment of the message may draw information from several segments of the sound (Fant, 1962). Moreover, despite the claim of distinctive feature theory that correlates of the features are to be found at every level of the speech process (articulatory, acoustic, auditory), Fant observed that "...statements of the acoustic correlates to distinctive features have been condensed to an extent where they retain merely a generalized abstraction insufficient as a basis for the quantitative operations needed for practical applications" (Fant, 1962, p. 94). He and his colleagues therefore set about describing the actual segments of sound observable in a spectrogram, the presumed elements with which the perceptual system has to work. They developed a terminology for describing these segments as manner or segment type features (e.g., voicing, plosive release, frication, nasal resonance, and so on) and place features specified in terms of formant pattern and energy distribution over frequency and time (Fant, 1968). On two general points, central to Fant's position, there now seems to be a growing consensus: First, one-to-one correspondence does not obtain between features of the signal and features of the message (cf. Parker, 1977; Ganong, Note 1); second, the signal takes effect through its entire spectral array rather than through the pattern of individual formants. However, Fant's approach has not yielded a solution to the segmentation puzzle.

An emphasis on the entire spectral array has also characterized the recent work of Stevens (e.g., 1972, 1975). Stevens has confronted the problem head-on by undertaking to reformulate the acoustic description of the signal to make it isomorphic with the message (that is, with the distinctive feature). He has adopted an explicitly evolutionary approach to the presumed link between production and perception. According to Stevens' (1972) quantal theory, phonetic categories have come to occupy those acoustic spaces where (by calculations from a vocal tract model) large articulatory variations have little acoustic effect, and to be bounded by those regions where small articulatory changes have a large acoustic effect. As a simple example, consider the articulatory-acoustic series that carries the speaker-listener from a high front vowel [i], through an alveolar fricative [s], to an alveolar stop [d]. However, most of Stevens' recent empirical work has concentrated on stop consonants. For example, Blumstein and Stevens (1979) derive three invariant spectral templates, determined by integration over a 26 msec window at the onset of the three voiced stops, /b,d,g/, before five representative vowels, spoken by two male speakers. Their terminology explicitly recalls

distinctive feature theory: "diffuse-rising" for alveolar, "diffuse-falling" for labial, and "compact" for velar. To match these acoustic properties, and as a step toward solving both the invariance and the segmentation problem, Stevens (1975) posits innate property detecting devices by means of which the infant is presumed to latch onto the speech system. We should note that these devices are conceived by Stevens as general to the mammalian auditory system rather than as tuned specifically to speech; in this they differ from the feature detectors to be discussed below.

Nonetheless, there are difficulties with Stevens' approach. First, the proposed stop-consonant templates invoke a fixed "locus," incompatible with the findings of Delattre, Liberman, and Cooper (1955). Thus, the supposedly invariant properties are static rather than dynamic and the role of their matching detectors is essentially to filter out "irrelevant" variation. Yet there is a mass of empirical data to demonstrate that the perceptual system is not only sensitive to the very information that these detectors are designed to exclude, but also uses it to reach phonetic decisions (see Studdert-Kennedy, 1976, 1980, for reviews). Second, the proposed property detectors are currently confined to consonant place of articulation, and there is reason to believe that an attempt to specify analogous detectors for, say, consonant voicing or vowel features would run foul of the many acoustic variations induced by phonetic context, rate and speaker. Finally, no integrative mechanism is proposed for aligning the features in the rows of the phonetic matrix within their appropriate segmental columns.

A third approach within this line of response has been that of researchers at Haskins Laboratories. They too have taken the goal of research in speech perception to be specification of relations between signal structure and linguistic units--but the units of traditional articulatory phonetics rather than of distinctive feature theory. They too early recognized the problem of acoustic segmentation, remarking at the conclusion of a study demonstrating contextual dependencies among plosive release bursts and following vowel: "...the irreducible acoustic stimulus is the sound pattern corresponding to the consonant-vowel syllable" (Liberman, Delattre, & Cooper, 1952, p. 516). However, they have been more atomistic in their approach, searching the signal for "minimal cues" (Liberman, Ingemann, Lisker, Delattre, & Cooper, 1959) to phonetic contrasts rather than characterizing the entire spectrum.

Two broad lines of research bearing on the segmentation issue have stemmed from this approach: One, deriving from a seemingly plausible account of categorical perception, has issued in recent work on feature detectors; the other, pursuing multiple acoustic cues, has sought to explain how they are integrated within the phonetic segment. Let us consider each in turn.

Early work with speech synthesizers showed that a useful procedure for defining the acoustic properties of a phoneme was to construct tokens of opponent categories, distinguished on a single phonetic feature, by varying a single acoustic parameter along a continuum (e.g., /ba/ to /da/, /da/ to /ta/, etc.). If listeners were asked to identify these tokens, they tended to identify any particular stimulus in the same way every time they heard it: There were few ambiguous tokens. Moreover, when asked to discriminate between neighboring tokens, listeners tended to do badly, if they assigned them to the

same phonetic category, well if they assigned them to different categories, even though the acoustic distance between tokens was identical in the two cases. This phenomenon was dubbed "categorical perception" (Liberman, Harris, Hoffman, & Griffith, 1957). Recent work has demonstrated that the phenomenon is not purely psychoacoustic, but, in some degree, a function of the listener's language (for a review, see Strange & Jenkins, 1978). Here, however, I want to pursue its possible psychoacoustic implications for segmentation--implications exploited, incidentally, by Stevens in his quantal theory (Stevens, 1972).

These implications have been elaborated (although with an eye to the problem of invariance rather than of segmentation) by recent work in selective adaptation. Following the work of Eimas and Corbit (1973), several dozen studies over the past five years have demonstrated that listeners, asked to identify tokens along a synthetic speech continuum before and after repeated exposure to (that is, adaptation with) a good category exemplar from one end of the continuum, report fewer instances from the adapting category after than before adaptation. Since this effect is observed on a labial voice onset time (VOT) continuum after adaptation with a syllable drawn from an alveolar continuum, and vice versa, adaptation is clearly neither of the syllable as a whole nor of the unanalyzed phoneme, but of a feature within the syllable. Eimas and Corbit therefore termed the adaptation "selective" and attributed their results to the "fatigue" of specialized detectors and to the relative "sensitization" of opponent detectors. Later studies have replicated the results for VOT and extended them to other featural continua, such as those for place and manner of articulation.

Unfortunately, several lines of evidence and argument undermine the hypothesis that selective adaptation reflects the operation of feature detecting mechanisms, specialized for speech. First is the fact that adaptation has been shown to depend on spectral overlap between adaptor and test continuum: Adaptation of consonantal features is specific to following vowel, syllable position and even fundamental frequency (see Ades, 1976, for a review), demonstrating that the supposed detectors are certainly not context-free (as would be required, if we wished to identify these mechanisms with Stevens' [1975] property detectors). Second, the grounds for arguing that, say, variations in voice onset time or place of articulation are mediated by opponent detectors, analogous to those posited for color perception, are not neurophysiological. On the contrary, the notion of binary opposition is drawn from distinctive feature theory, despite the fact that this theory holds such oppositions to be abstract relations within a phonological system, often realized phonetically by multiple, context-dependent values (cf. Parker, 1977). Finally, the proposed feature detectors lack biological warrant. The communication system of an animal reflects the pressures of its environment and life-history. Animals such as bullfrogs (Capranica, 1965) or certain songbirds (Marler, 1963) require innate feature detectors or templates, because they must identify their species accurately, and must learn to do so (when learning is involved) within a relatively brief (or non-existent) period of parental care. The message conveyed by their species-specific signals is simple and largely confined to matters associated with reproduction. This is not the case for human infants. The patterns of mother-infant interaction (e.g., Stern, Jaffe, Beebe, & Bennett, 1975; Freedle & Lewis, 1977) and of early infant vocalization (e.g., Huxley & Ingram, 1971, 162 ff.; Menn, 1979)

suggest a lengthy, epigenetically guided search for sound structure rather than an innate response to species-specific calls.

I take the preceding arguments and evidence to rule out feature or property detecting mechanisms specialized for speech, but to say nothing about property detecting mechanisms in the general auditory system. In fact, selective adaptation studies bear on the segmentation issue precisely by reaffirming the familiar fact that listeners can engage distinct channels of analysis to perceive contrasts between properties of acoustic signals and that, in speech, these properties may be the many-to-one or one-to-many correlates of phonetic features.

Whether or how listeners normally use these channels in listening to speech is both unknown and a separate issue, one that returns us to the second line of research that has developed out of the Haskins work—the search for cues. This work can be seen as, in a sense, coordinate with Stevens' research on spectral properties. Just as Stevens' work raises the question of how diverse spectral properties, said to correspond to distinctive features, are organized and aligned into phonemes and syllables, so the Haskins work raises the issue of how diverse acoustic cues are integrated perceptually into spectral and temporal properties corresponding to features or phonemes.

The puzzle arises because each phonetic distinction is susceptible to manipulation by many acoustic cues. For example, Lisker (1978) has listed sixteen different cues (not all of them tested, it is true) that may serve to distinguish the medial stops of rapid and rabid. Similarly, Bailey and Summerfield (1978) have shown that perceived place of articulation of a stop consonant (/p/, /t/ or /k/), consequent upon the introduction of a brief silence between /s/ and a following vowel, depends in English on the duration of the silent closure, on spectral properties at the offset of /s/ and on the relation between these properties and the following vowel.

Typically, cues seem to form a hierarchy with one or more cues dominating the others. Thus, presence of voicing during medial closure forces perception of rabid rather than rapid, but its absence permits other cues to come into play. These cues may then engage in trading relations, so that equivalent percepts result from reciprocal increases and decreases in their values. A well-worked example is provided by the distinction between slit and split. Here, equivalent percepts of split are yielded either by a relatively brief silence (stop closure) after /s/, followed by second and third formant transitions appropriate to /p/, or by a longer silence followed by steady-state vowel formants without transitions (Fitch, Halwes, Erickson, & Liberman, in press). (For a review of such work, see Liberman & Studdert-Kennedy, 1978.)

At first glance, multiple cue equivalences, or trading relations, seem commonplace in light of the familiar intensity-time relations of audition and vision. However, while constant energy functions are readily rationalized in terms of basilar mechanics or retinal photochemistry, there is no obvious account of why cues as diverse as, say, silence and formant transitions should be perceptually equivalent. No less puzzling is the function of such spectrally diverse and temporally distributed cues in normal perception. Given the multiplicity of cues to any given phonetic distinction, it hardly

seems plausible that each cue be extracted by a separate channel and then combined with other cues to yield a phonetic feature--which must then be combined with other phonetic features to form a phone or syllable (Pisoni & Sawusch, 1975). What principle would rationalize these successive integrations?

The quandary was recognized and a rationale for its solution proposed a number of years ago by Lisker and Abramson (1964, 1971). They pointed out that the diverse array of cues which separate so-called voiced and voiceless initial stop consonants in many languages--plosive release energy, aspiration energy, first formant onset frequency--were all consequences of variations in timing of the onset of laryngeal vibration with respect to plosive release--in other words, of voice onset time (VOT). Moreover, they proposed VOT as no more than an instance of a general articulatory variable, timing of laryngeal action, from which the multiple acoustic cues to consonant voicing in all contexts (initial, medial, final/stressed, unstressed) might be derived (Abramson, 1977; cf. Fant, 1960, p. 225). Unfortunately, VOT has often been narrowly interpreted as an acoustic variable, comparable with supposed non-speech analogs, such as the relative onset time of noise-buzz sequences (Stevens & Klatt, 1974; Miller, Wier, Pastore, Kelly, & Dooling, 1976) or of two tones (Pisoni, 1977). Such studies have diverted attention from the deeper issue that Lisker and Abramson were addressing, namely, the origins of acoustic cue diversity.

If we extend their account to perception, we have to say that the perceptual counterpart of the unitary articulatory gesture is not an arbitrary collection of cues, but an integral auditory array. In apprehending this array, we perceive the event that shaped it. In other words, we perceive the gesture by means of its radiated sound pattern, just as we perceive the movement of a hand by reflected light--or the articulated gestures of speech by cineradiography. This is not a new notion. Both Paget in his book on human speech (1930, chap. VII) and Dudley (1940) some years later pointed out that the sounds of speech could be regarded as the carriers of articulated gesture.

Yet, attractive as this view may be and important as an account of the origins of speech cue diversity, it still does not explain how we divide the gesture into its underlying linguistic segments. On the contrary, if we regard the consonant-vowel syllable as the product of an integrated, ballistic gesture (Stetson, 1952) and if, further, we regard this gesture as the essential perceptual object that underlies the diverse acoustic cues, we are led to the paradoxical conclusion that the atomistic approach of the Haskins researchers returns us to a view that anchors perception in the entire spectral array rather than in individual cues--a view, moreover, that comes close to that of the more pragmatically directed line of response to the segmentation problem, mentioned at the beginning. This approach makes no attempt to align segments of the signal with the abstract phonetic segments that shape it.

Let us turn briefly to this second line of response. In the first instance, the approach was adopted in order to synthesize or compile speech with "building blocks" (Harris, 1953), formed from half syllables or "phoneme dyads" (Peterson, Wang, & Sivertsen, 1958). Given our ignorance of precisely

how the consonant gesture merges with the vowel gesture to yield the motoric consonant-vowel syllable, this was an eminently practical approach to acoustic synthesis, and one that is still viable (e.g., Fujimura, 1975; Mattingly, 1976). Recently, Klatt (1980) has developed a detailed model in which elements larger than the phone are also the elements of perception (cf. Fujimura, 1975). Klatt proposes a store of a few hundred spectral templates corresponding to phone-pairs, as sufficient to bypass application of phonological rules during perception and to enable recognition of a sizeable lexicon. The point of interest to the present discussion is that Klatt's phone-pairs are, in principle, no different than the acoustic counterparts of syllable gestures. He has taken seriously the familiar claim that "...phones are not directly perceived, but must rather be derived from a running analysis of the signal over stretches of at least syllable length" (Liberman & Studdert-Kennedy, 1978, p. 153), but has elected to sidestep their analytic derivation. What we must ask is whether this brute force solution is our only recourse.

I believe that it is not, and that the solution to the problem lies in recognizing that the speech signal can indeed be segmented into acoustic groups corresponding to phonetic segments, but only by an organism that already knows that phonetic segments are there to be found. The point is clearly made by the results of recent work on reading spectrograms (Cole, Rudnicky, Reddy, & Zue, 1980). The subject, VZ, is a skilled acoustic phonetician who has devoted some 2500-3000 hours to learning to read spectrograms. What is of interest here is that while VZ's performance on correctly labeling segments seems to hover around 85% (a vast improvement, incidentally, on previous reported work), he identifies the existence of segments with an accuracy close to 97%. What is the basis of this remarkable performance?

A moment's reflection tells us that the performance must rest on two crucial facts: First, VZ knows that the segments are there to be found; second, the spectrographic display represents sufficient acoustic information for the task to be accomplished. Consider each in turn.

That VZ's skill rests on his knowledge that phonetic segments exist becomes obvious as soon as we imagine a reader who lacks this knowledge and confronts a spectrogram as a cryptoanalyst, knowing nothing more than that it conveys a message. How would he proceed? Let us grant that, being human, the cryptographer would start by looking for units (very much like the epigraphist, confronted with Minoan Linear B [Chadwick, 1958]). What he would find would be, of course, what Fant (1968) found, namely, a large number of clearly defined acoustic segments bearing (as we know, but the cryptographer does not) a one-to-many and many-to-one relation to the phonetic message. Since this appears to be precisely the condition of the human infant, we must ask how the infant acquires the knowledge that segments exist.

Before attempting to answer this, consider the second condition of VZ's performance--that the spectrographic display does represent enough information for the task to be done. What information does VZ, in fact, use? Apparently, the primary sources of information are spectral discontinuities (including, we may assume, the brief silences of stop closure, contrasts between friction noise and vowel periodicity, formant transitions, nasal resonances, plosive release bursts, and so on) and duration. These are the properties recommended

by Fant (1968) in his prescription for spectrogram reading. They are also the properties described by Bondarko (1969) in her account of within-syllable segmentation as based on auditory contrast, a relational process, rather than on the extraction of absolute, context-free features.

Yet, as we have already argued, use of this contrast for segment recognition rests on the prior knowledge that phonetic segments exist. Only when the cryptographer possesses this key to the code, is he in a position to discover, by prolonged practice, the groupings and divisions of the acoustic stream that relate the signal to its phonetic message. How then does the human infant learn the code?

Perhaps the answer will emerge from a deeper understanding of the development of imitation. The process is epigenetic: It seems to rest on an innate, imitative response to the sounds of speech, gradually shaped by the language to which the infant is exposed. The newborn quickly learns to discriminate sound from silence (Friedlander, 1970), voices from other sounds (Alegria & Noirot, Note 2), its mother's voice from a stranger's, intonation from monotone (Mehler, Bertoncini, Barrière, & Jassik-Gershenfeld, 1978), and, in due course, syllable from intonation contour. Motorically, within weeks of birth, the infant is able to imitate the facial movements of its caretakers (Meltzoff & Moore, 1977), and soon engages in "prespeech" oral play, watching its mother's eyes and mimicking the movements of her mouth (Trevarthen et al., 1975). Gradually, it shifts from cooing and crowing to intonated babble, until, before the end of its first year, syllables emerge.

Here, the process of differentiation ends--saving us from the paradoxical claim that the infant can imitate what it cannot perceive: The babbling infant imitates syllables, not phonetic segments. For, as we have seen, phonetic segments exist neither in the articulatory gesture nor (a fortiori) in the acoustic signal. Rather, phonetic segments are abstract control processes that emerge as the links between acoustic syllables and their corresponding gestures.

We do not have to suppose that development requires actual activation of the motor system, although this must surely facilitate the process. That activation is not necessary is shown by several well-attested cases of children who have learned to understand without being able to speak. An impressive case is that of Richard Boydell, a victim of congenital cerebral palsy, who, unable to speak and lacking all use of hands and arms, nonetheless learned by prolonged maternal tutelage to understand speech, to read, and even, when provided at the age of 30 with a foot-typewriter, to express his thoughts in highly literate English (Fourcin, 1975). If the difficulties of purely auditory segmentation are indeed as real as the evidence suggests, we must conclude that what remained unimpaired in Boydell was the sensori-motor center that links sound and gesture, and that permits the imitating infant to discover the abstract components of its language.

If this view has any merit and any import for future research, it lies in the implication that we cannot understand speech perception by simply charting relations between signal and percept. We have to go behind the signal to the gestures that shape the resonant volumes of the vocal tract. We have to understand how consonant movement and vowel movement merge to form the motor

syllable. Perhaps we shall then conclude that the sounds of speech do indeed form a natural class of dynamic events--distinctive in that those who perceive them have learned how to speak.

REFERENCE NOTES

1. Ganong, W. F. The internal structure of consonants in speech perception: Acoustic cues, not distinctive features. Unpublished manuscript.
2. Alegria, J., & Noirot, E. Neonate orientation behavior towards the human voice. Paper read before XVth International Congress of Ethology, Bielefeld, Germany, August 22-26, 1977.

REFERENCES

Abramson, A. S. Laryngeal timing in consonant distinctions. Phonetica, 1977, 34, 295-303.

Ades, A. E. Adapting the property detectors for speech perception. In R. J. Wales & E. Walker (Eds.), New approaches to language mechanisms. Amsterdam: North-Holland, 1976.

Bailey, P. J., & Summerfield, Q. Some observations on the perception of [s] + stop clusters. Haskins Laboratories Status Report on Speech Research, 1978, SR-53(2), 25-60.

Blumstein, S. E., & Stevens, K. N. Acoustic invariance in speech production. Journal of the Acoustical Society of America, 1979, 66, 1001-1018.

Bondarko, L. V. The syllable structure of speech and distinctive features of phonemes. Phonetica, 1969, 20, 1-40.

Capranica, R. R. The evoked vocal response of the bullfrog. Cambridge, Mass.: M.I.T. Press, 1965.

Chadwick, J. The decipherment of Linear B. Cambridge, England: Cambridge University Press, 1958.

Cole, R. A., Rudnicky, A., Reddy, R., & Zue, V. W. In R. A. Cole (Ed.), Perception and production of fluent speech. Hillsdale, N.J.: Erlbaum, 1980.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America, 1955, 27, 769-773.

Dudley, H. The carrier nature of speech. The Bell System Technical Journal, 1940, 19, 495-515. (Reprinted in J. L. Flanagan & L. R. Rabiner (Eds.), Speech synthesis. Stroudsburg, Pa.: Dowden, Hutchinson and Ross, 1973, 22-42.)

Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. Cognitive Psychology, 1973, 4, 99-109.

Fant, C. G. M. Acoustic theory of speech production. The Hague: Mouton, 1960.

Fant, C. G. M. Descriptive analysis of the acoustic aspects of speech. Logos, 1962, 5, 3-17.

Fant, C. G. M. Analysis and synthesis of speech processes. In B. Malmberg (Ed.), Manual of phonetics. Amsterdam: North Holland, 1968, 173-277.

Fitch, H. L., Halwes, T., Erickson, D., & Liberman, A. M. The perceptual equivalence of trading-relation cues. Perception & Psychophysics, in press.

Fourcin, A. J. Language development in the absence of expressive speech. In E. H. Lenneberg & E. Lenneberg (Eds.), Foundations of language

development (Vol. 2). New York: Academic Press, 1975, 263-268.
Freedle, R., & Lewis, M. Prelinguistic conversations. In M. Lewis & L. A. Rosenblum (Eds.), Interaction, conversation and the development of language. New York: Wiley, 1977, 157-186.
Friedlander, B. A. Receptive language development in infancy. Merritt-Palmer Quarterly, 1970, 16, 7-51.
Fujimura, O. Syllable as a unit of speech recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, 23, 82-87.
Harris, C. M. A study of the building blocks of speech. Journal of the Acoustical Society of America, 1953, 25, 962-969.
Hockett, C. F. A course in modern linguistics. New York: MacMillan, 1958.
Huxley, R., & Ingram, E. (Eds.). Language acquisition: Models and methods. New York: Academic Press, 1971.
Klatt, D. H. Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), Perception and production of fluent speech. Hillsdale, N.J.: Erlbaum, 1980.
Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII; Perception. New York: Springer-Verlag, 1978, 143-178.
Liberman, A. M., Delattre, P. C., & Cooper, F. S. The role of selected stimulus variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 1952, 65, 497-516.
Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 1957, 53, 358-368.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.
Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P. C., & Cooper, F. S. Minimal rules for synthesizing speech. Journal of the Acoustical Society of America, 1959, 31, 1490-1499.
Lisker, L. Rapid vs rabid: A catalogue of acoustic features that may cue the distinction. Haskins Laboratories Status Report on Speech Research, 1978, SR-54, 127-132.
Lisker, L., & Abramson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.
Lisker, L., & Abramson, A. S. Distinctive features and laryngeal control. Language, 1971, 47, 767-785.
Marler, P. Inheritance and learning in the development of animal vocalizations. In R.-G. Busnel (Ed.), Acoustic behavior of animals. Amsterdam: Elsevier, 1963, 228-243 and 794-797 (addendum).
Marslen-Wilson, W. D., & Welsh, A. Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 1978, 10, 29-63.
Mattingly, I. G. Syllable synthesis. Journal of the Acoustical Society of America, 1976, 60, 575. (Abstract)
Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gershenfeld, D. Infant recognition of mother's voice. Perception, 1978, 7, 491-497.
Meltzoff, A. N., & Moore, M. K. Imitation of facial and manual gestures by human neonates. Science, 1977, 198, 75-78.
Menn, L. Pattern, control and contrast in beginning speech. Bloomington: Indiana University Linguistics Club, 1979.
Miller, J. D., Wier, C. C., Pastore, R., Kelly, W. J., & Dooling, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-

lead times: An example of categorical perception. Journal of the Acoustical Society of America, 1976, 60, 410-417.

Nakatani, L. H., & Dukes, K. D. Locus of segmental cues for word juncture. Journal of the Acoustical Society of America, 1977, 62, 714-719.

Nottebohm, F. Asymmetries of neural control of vocalization in the canary. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.), Lateralization in the nervous system. New York: Academic Press, 1977, 23-44.

Paget, R. Human speech. London: Routledge and Kegan Paul, 1930.

Parker, F. Distinctive features and acoustic cues. Journal of the Acoustical Society of America, 1977, 62, 1051-1054.

Peterson, G. E., Wang, W. S.-Y., & Sivertsen, E. Segmentation techniques of speech synthesis. Journal of the Acoustical Society of America, 1958, 30, 739-742.

Pisoni, D. B. Identification and discrimination of the relative onset times of two component tones: Implications for the perception of voicing in stops. Journal of the Acoustical Society of America, 1977, 61, 1352-1361.

Pisoni, D. B., & Sawusch, J. R. Some stages of processing in speech perception. In A. Cohen & S. G. Nooteboom (Eds.), Structure and process in speech perception. New York: Springer-Verlag, 1975, 16-35.

Stern, D. N., Jaffe, J., Beebe, B., & Bennett, S. L. Vocalizing in unison and in alternation: Two modes of communication within the mother-infant dyad. In D. Aaronson & R. W. Rieber (Eds.), Developmental psycholinguistics and communication disorders. New York: New York Academy of Sciences, 1975, 89-100.

Stetson, R. H. Motor phonetics. Amsterdam: North Holland, 1952.

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw Hill, 1972, 51-66.

Stevens, K. N. The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech. New York: Academic Press, 1975, 303-330.

Stevens, K. N., & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 1974, 55, 653-659.

Strange, W., & Jenkins, J. J. The role of linguistic experience in the perception of speech. In H. L. Pick, Jr. & R. D. Walk (Eds.), Perception and experience. New York: Plenum, 1978.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), Contemporary issues in experimental phonetics. New York: Academic Press, 1976, 243-293.

Studdert-Kennedy, M. Speech perception. Language and Speech, 1980, 23, 45-65.

Trevarthen, C., Hubley, P., & Sheeran, L. Les activités innées du nourisson. La Recherche, 1975, 56, 447-458.

van Lawick-Goodall, J. In the shadow of man. London: Collins, 1971.

Wilson, E. O. Sociobiology: The new synthesis. Cambridge, Mass.: Belknap Press, 1975.

READING, LINGUISTIC AWARENESS AND LANGUAGE ACQUISITION*

Ignatius G. Mattingly+

INTRODUCTION

Most of us who speculate about the reading process begin by considering the nature of the relationship between listening to speech and reading. Are these two cognitive processes essentially the same, apart from a difference in input modality? Or are they essentially quite different, despite their shared linguistic character? My view is that reading, though closely related to listening, is different from it in some very crucial respects.

In an earlier paper (Mattingly, 1972), I attempted to characterize the difference in terms of "primary" and "secondary" linguistic activity. I suggested that, while the primary linguistic activities of speaking and listening are natural in all normal human beings, secondary linguistic activities, such as versification and reading, are parasitic on these primary activities, and require "linguistic awareness," a specially cultivated meta-linguistic consciousness of certain aspects of primary linguistic activity. I still believe this distinction to be a valid one, but I now think that linguistic awareness is not a matter of consciousness, but of access. This access is probably largely unconscious, but the degree of consciousness is not very relevant. Moreover, what the linguistically aware person has access to is not his linguistic activity--the processes by which he produces and understands sentences--but rather, his knowledge of the grammatical structure of sentences. Finally, I would not now wish to imply that secondary activity is less natural than primary linguistic activity. I will argue, in fact, that reading involves not only the mechanisms of speech understanding but also those of language acquisition, and that it is just as natural, and in a sense more "linguistic," than listening to speech.

It may perhaps disarm criticism to some degree if, before proceeding further, I distinguish two modes of mental activity that might be called reading. In the first mode, the reader identifies written words in a sentence as corresponding to specific items in his mental lexicon and makes a grammatical analysis, as a result of which he may be said to "understand" the sentence. This mode might be called "analytic" reading. In the second mode

---

*Paper presented at the International Reading Association Seminar, Linguistic Awareness and Learning to Read, Victoria, B.C., June, 1979. To appear in the proceedings of the seminar, J. Downing (Ed.), untitled. Some of the material in this paper has also appeared in somewhat different form in Mattingly (1978) and Liberman, Liberman, Mattingly, and Shankweiler (1978).

+Also University of Connecticut.

Acknowledgment: Support by the International Reading Association and by Grant HD-01994 from the National Institute of Child Health and Human Development is gratefully acknowledged.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

of reading, which might be called "impressionistic" reading, the reader tries to guess the meaning of the text just by looking at the words, without making specific lexical identifications and without making a grammatical analysis. This mode of reading relies on the fact that a written word, just because it is a familiar orthographic pattern, and not because it corresponds to a lexical item, is capable of evoking a rich network of semantic associations. It would not be surprising if such evocation were shown to occur much more rapidly than the identification of a word as a specific lexical item.

In what follows, I am concerned almost entirely with analytic reading, justifiably, I feel. It may well be that, relying on the semantic associations of orthographic patterns and on a priori knowledge, a reasonably intelligent impressionistic reader can get the general sense of a text. Analytic reading may be slower and more laborious than impressionistic reading. Analytic reading may even be a relatively rare act on the part of a skilled reader; depending on the nature of the text and his motivation in reading it, he may be reading impressionistically most of the time. Yet I believe that, useful as it may be to be able to read impressionistically, a person is not a reader if he cannot read a sentence analytically when it is really essential to the understanding of a text to do so.

SOME LINGUISTIC AND PSYCHOLINGUISTIC ASSUMPTIONS

According to the generative linguists, the knowledge that an ideal speaker-hearer has about the structure of his language is mentally represented in a grammar. The grammar may be viewed as a device for specifying the linguistically relevant aspects of any and all sentences in the language. It consists of syntactic, phonological and semantic components, each of which is a set of ordered rules; and a lexical component, each entry in which specifies the peculiar syntactic, phonological and semantic properties of a word in the language. The phrase-structure rules and transformational rules of the syntactic component "generate," i.e., derive, the phrase-marker that represents the syntactic structure of a sentence. Rules of lexical insertion (also part of the syntactic component) relate the words of the sentence to lexical entries. The rules of the phonological component generate the phonetic representation--the intended or perceived pronunciation of the sentence--given the phrase-marker and the lexically-specified phonological properties of each word. Analogously, the rules of the semantic component generate the semantic representation--the meaning of the sentence--given the phrase-marker and the lexically specified semantic properties of each word. The phrase marker, the lexical content of the sentence, and the phonetic and semantic representations constitute the "surface-structure" description of the sentence (Chomsky, 1975, 1977).

An actual speaker-hearer's "grammatical knowledge" is no doubt very imperfect. It is also tacit knowledge: The speaker-hearer knows the grammar of his language, but need not "know that he knows" it, or be able to formulate it coherently. Yet grammatical knowledge is accessible, in the sense that the speaker-hearer has intuitions about grammaticality. He is able to say whether a certain phonetic contrast is distinctive in his language, whether a certain syntactic pattern is acceptable, whether a certain sentence is meaningful. The validity of these intuitions is corroborated by the success of linguists in reconstructing descriptively adequate grammars. But there are limitations

on the scope of grammatical knowledge. The speaker-hearer has very limited intuitions, for example, about the acoustic properties of the speech signal that can be shown to determine his phonetic perceptions (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Accordingly, the grammar has nothing to say about the complex relationships between the phonetic representation of a sentence and its acoustic realization.

A child acquiring the grammar of his native language is rather in the position of a linguist (Chomsky, 1965). Given a theory of language specifying the structural properties that all grammars share, and data as to correspondences between sound and meaning, he proceeds to construct the lexicon and the grammatical rules. The child's position is different from that of the linguist mainly in that his general theory of language is innately given, and superior to any general theory so far formulated by linguists. But having a task similar to the linguist's, he must have psychological mechanisms for doing what linguists do: making hypotheses about rules and about the content of lexical entries, constructing hypothetical phonetic and semantic representations, and comparing them with the available data.

The relationship of grammatical knowledge to actual speaking and listening in real time appears to be rather indirect. The task of the speaker is to determine (and realize articulatorily) the phonetic representation of a sentence, starting with information (his knowledge, motives, intentions) that most directly constrains the semantic representation; the task of the listener is to determine the semantic representation, starting with information (the auditory properties of the acoustic signal) that most directly constrains the phonetic representation. It is rather unlikely that either speaker or listener obtains the required representation by generating it, or that either obtains intermediate representations by applying generative rules in reverse order. Instead, various analytic mechanisms--mechanisms for speech perception, phonological analysis, lexical search, semantic analysis--seem to be in operation. (See the discussion of these questions in Fodor, Bever, and Garrett, 1974, Chap. VI.) These performance mechanisms are heuristic in character, apparently using pragmatic, obviously fallible strategies. For example, a good parsing strategy for English appears to be: Assume that the elements of any sequence of the form NP V(NP) are the subject, verb and object of the same clause. (Fodor et al., 1974, p. 345). A plausible strategy for lexical search might well be to list for each item in the lexicon its possible phonetic realizations in various phonological contexts, thus avoiding phonological analysis.

While it is useful to postulate the existence of specific analytic mechanisms with particular functions, it is probably incorrect to envision these mechanisms as a series of modules, the output of each becoming the input to the next. To account for our perceptions, such a model would require provisions for feedback of information from later mechanisms to earlier ones, and the more feedback is assumed, the more arbitrary the functional separation of the different mechanisms appears. It seems more prudent to make the weaker assumption: that the analytic mechanisms somehow collaborate in concurrently reconstructing the various parts of the surface-structure description, given an input that is itself of mixed character. Thus the reconstruction of the phonetic representation in listening is based not only on acoustic information but on semantic presuppositions and hypotheses, and the semantic, syntactic

and lexical mechanisms play a part as well as the speech perception mechanism. In the process, other parts of the surface-structure description are partially determined as well.

The sets of strategies used by the various mechanisms might be referred to as "performance knowledge." Performance knowledge, unlike grammatical knowledge, seems to be relatively inaccessible. The speaker-hearer does not have intuitions about performance. What is known about performance mechanisms--the speech perception mechanism or the parsing mechanism, for example--has been learned by experimental inference rather than by linguistic analysis.

Though performance knowledge is not the same thing as grammatical knowledge, the two must somehow be related. Which strategies it is appropriate to try in what order obviously depends upon the grammar of the language, even though the performance mechanisms are not doing grammatical derivations. It has been suggested that the capacity to acquire language includes the ability to compute the optimal set of analytic strategies for a given grammar (Fodor et al., 1974, p. 372). Thus, as grammatical knowledge develops during language acquisition, performance knowledge would increase as well.

How much must an actual speaker-hearer know about the grammar of his language to insure a degree of performance knowledge sufficient for ordinary speaking and understanding? Perhaps relatively little, in comparison with the ideal speaker-hearer of linguistic theory. It is quite believable, for example, that a person might have parsing strategies that could cope, much of the time, with passive constructions, without having grammatical knowledge of the rules for generating passive sentences.

To put the matter somewhat differently, the grammatical knowledge a language-learner is potentially capable of acquiring far exceeds the functional requirements of performance. But if this is so, we should not find it surprising that some speaker-hearers, driven by an instinctive linguistic curiosity, continue acquiring the grammar of their language indefinitely, while others essentially abandon language acquisition once the performance mechanisms are adequately equipped for the purposes of ordinary communication.

If one takes seriously the conception of the infant language-learner as linguist, one might suppose that during the period of active language acquisition, grammatical knowledge would be highly accessible (access, to repeat, does not imply consciousness). But if grammatical knowledge is not directly used in linguistic performance, it is to be expected that after language acquisition has ceased to be a major preoccupation, grammatical knowledge should tend to become less accessible.

ORTHOGRAPHY AND READING

To clarify the relationship of reading to speaking and listening, I will discuss two kinds of evidence. The first kind of evidence derives from consideration of practical orthographies. The second kind of evidence is experimental: Performance in tasks that are similar to reading can be compared with performance in tasks involving production or perception of spoken language.

What aspects of a sentence do practical orthographies transcribe? The apparent heterogeneity of orthographies might suggest that there is no one answer to this question; the traditional classification of orthographies into logographic, syllabary and alphabetic modes seems to imply that each mode transcribes sentences in a different way. Yet if there is not some clear sense in which all orthographic modes are alike, the unappealing possibility that a separate account of the reading process must be given for each mode will have to be entertained.

I begin with the observation that the written form of a sentence, in any practical orthography, stands for its lexical content, for the words of the sentence. If this point seems unnecessarily obvious, consider that, in principle, non-lexical orthographies are quite possible. An orthography based on speech articulations, in which the symbols represent vocal-tract shapes, was indeed proposed by Wilkins (1668) and later by Bell (1867). An orthography based on acoustic waveforms, or better, on patterns like Haskins Laboratories' idealized spectrograms, is also conceivable, and obviously parallels the acoustic input to the listener's analytic mechanisms. Though it will be maintained below that practical orthographies are not, in fact, phonetic, it would be possible to have an orthography based on the phonetic representation, essentially a narrow phonetic transcription. The most plausible orthography of all, perhaps, would be one based on semantic representations of sentences, since the semantic representation is what the writer wants to convey to the reader. Perhaps the notational system of mathematics, logic and the sciences might be regarded as specialized semantic orthographies.

A further limitation on the character of orthographies is that they do not transcribe information about the phrase-marker that represents the syntactic structure of the sentence. This is of special interest because, in the grammar, if the phrase-marker and the lexical content of the sentence are given, the other parts of the surface-structure description--the semantic and phonetic representations--can be derived. Orthographies transcribe words, but not the syntactic information that complements the information conveyed by the words.

The second generalization to be made about practical orthographies is that the lexical items are transcribed morphemically. In the lexical component of the grammar, phonological information about the word appears in a morphophonemic representation (Chomsky & Halle, 1968). For example, in the lexicon of English, the words heal, health, healthy have the representations /hēl/, /hēl+θ/, /hēl+θ+y/; /hel/, /θ/ and /y/ being morphemes (the phonological symbols used here are convenient abbreviations for sets of distinctive-feature values, and "+" indicates a morpheme boundary). A morpheme has semantic as well as phonological value, but of course the semantic value of a word is not necessarily predictable from the semantic values of its component morphemes.

It is obviously possible to transcribe this representation in a number of equivalent ways: using a distinct symbol for each morpheme (the logographic method), or for each syllable (the syllabary method) or for each morphophoneme (the alphabetic method), or even for each distinctive-feature value (too cumbersome an approach for a practical orthography). But practical orthographies are morphemic in the sense that, in general, a morpheme is transcribed

in the same way wherever it occurs.

The morphemic character of logographic systems, such as the one used for Chinese and borrowed by the Japanese, is quite obvious, since a separate character is used for each morpheme of a word. (It is perhaps worth insisting on the distinction between words and morphemes in this connection. Modern Chinese has many one-morpheme words but also a great many compounds. Thus a reader who knows the characters for only a few thousand common morphemes is able to read many thousands of words.)

Since a morpheme is a pairing of semantic and phonological values, it is not surprising to find that the most common kind of Chinese character consists of a "radical" or semantic element, itself a character standing for a morpheme of related meaning, and a "phonetic" element, a character standing for a morpheme that is, or once was, phonologically similar (Martin, 1972). It is also clear that in this logographic system, the characters stand for morphemes as such, and not for sequences of morphophonemes, because a character never corresponds to an arbitrary sequence of morphophonemes, and because homophonous morphemes are regularly assigned distinct characters. On the other hand, the fact that morphemes, which are meaning-bearing elements, are so obviously the units of transcription does not compel the conclusion that Chinese and Japanese readers must "go directly to meaning." The point is rather that words of a sentence are transcribed according to the morphemic structure of their lexical entries.

The essentially morphemic character of alphabetic and syllabary systems is perhaps less obvious. In the first place, it has to be shown that an alphabetic orthography, such as that of English, is not, as sometimes assumed, a phonetic orthography.

In the grammar, the phonetic representation is generated by the application of phonological rules to the morphophonemic forms in the lexicon. Thus the rules shorten the long vowel of /hēl+θ/ ("Laxing"), yielding [helθ], and they add a following glide ("Diphthongization") and shift the quality of the same vowel in /hēl/ ("Vowel Shift"), yielding [hīyl]. Other rules assign varying degrees of stress to both words, depending upon their position and syntactic function in the sentence. Similarly, through the application of the relevant phonological rules, morphophonemic /tele+græf/, /tele+græf+ik/, /tele+græf+y/ yield phonetic [tèlәgræf], [tèlәgrǽfik], [tәlégrәfiy]. Thus, in the phonetic representation, an underlying morpheme is not consistently represented as it is in the morphophonemic representation (Chomsky & Halle, 1968). Clearly, as Chomsky has argued (1964), the conventional spellings of these English words correspond to the morphophonemic rather than the phonetic forms.

The morphophonemic character of an alphabetic orthography is of course more obvious in the case of a language with a relatively "deep" phonology, such as English or French. The orthography of a language with a shallow phonology will inevitably be fairly close to the phonetic representation, since the morphophonemic representation itself is close to the phonetic representation. This seems to be true for the orthographies of Finnish, Vietnamese and Serbo-Croatian, for example, which are often said to be "phonetic," but are not really exceptions.

For some languages with simple syllable structure, syllabary systems are used, but it is still the case that the transcription is at the morphophonemic level. For example, it is a rule of Japanese phonology that a non-initial voiceless stop becomes voiced (Martin, 1972). In the Romanized forms kana, hiragana, the effect of this rule is explicit: The initial /k/ of /kana/ in the one-morpheme word becomes [g] in the compound. But in the hiragana syllabary system that is one way of transcribing Japanese, the same kana character is used for the syllable /ka/ in both words.

Yet it must be admitted that some orthographies, though not phonetic, are yet less than perfectly morphophonemic. In Turkish, the alternations determined by the Vowel Harmony rule are transcribed, perhaps because there are numerous borrowed words not subject to this law (A. Kardestuncer, personal communication). In Spanish, infinitives are transcribed without the phonologically deleted final /e/ of the morphophonemic representation; thus /decire/, 'to say' is written decir (Harris, 1969). In Sanskrit, the alternations between aspirated and inaspirated stops (Grassman's Law) are transcribed.

Granted that alphabetic systems are morphophonemic (though imperfectly so), it is now argued that they are morphophonemic in order to represent morphemes consistently, rather than to represent the morphophonemes as such. If alphabetic orthographies were entirely systematic, it might not be possible to demonstrate this convincingly. But in the case of English the inconsistencies can be turned to account. Thus it is quite easy to demonstrate that in English a morphophoneme may be spelled in a number of different ways, and that, worse still, the spellings of one morphophoneme overlap with those of another. For example, morphophonemic /e/ is spelled ee or ie or ea or eCe or iCe; ie and iCe can also spell /i/ and ea can spell /e/ and /æ/. But notice that these variations often serve to distinguish homophonous morphemes from one another, as in sea, see; meet, meat, mete; and that despite all these inconsistencies, a particular morpheme is generally spelled in the same way in its occurrences in different words. Thus /hēl/ is heal in heal, healthy, healthful. There are obviously some exceptions; it is too bad that fashion, delusion, cylinder are not spelled *facion, *deludion, *cylindr, respectively (see Klima, 1972, for discussion); and of course English orthography makes no attempt to cope with genuine morphological irregularities: the past tense of /θink/ is written thought, not *thinked.

In the case of syllabary systems the point can be made in a different way. A syllabary is preferable either to an alphabetic or to a logographic system from the standpoint of learnability and convenience. But since the alphabetic principle became well known, syllabaries have not been widely used. It seems to be a desirable if not essential condition for using a syllabary not only that the syllable structure of the language should be simple, but also that morpheme boundaries should coincide with syllable boundaries. If this condition is not met--and it is not in many Indo-European languages--morphemes cannot be consistently transcribed. The limited use of syllabaries thus probably attests to the importance of the morpheme.

In sum, a practical orthography conveys the lexical content of a sentence by transcribing the words morphemically. Differences in orthography reduce to whether a morpheme is written as a single symbol or as a sequence of symbols corresponding to morphophonemes or morphophonemic syllables.

This characterization of practical orthographies suggests that in the actual process of reading, the analysis of a sentence begins with its lexical content and not with its phonetic representation; I return to this point later. It suggests also that lexical items are recognized by virtue of their morphological or (in the case of alphabets and syllabaries) their morphophonemic structure. But it would be a mistake to conclude that such structure is the only basis for word recognition. The semantic associations of an orthographic form (the basis for what I have called impressionistic reading) apparently tell the reader very quickly that the word is one he has seen before. If the word is very familiar, they are also sufficient for lexical identification. It must be assumed that Chinese characters with no internal structure, and one-letter words in English, at least, are identified in this way. The effect is enhanced if the orthographic form is "glyphic," i.e., compact and visually distinct (Brooks, 1977). On the other hand, if the word is one that the reader has never seen before, though it is part of his spoken vocabulary, it is obvious that the morphophonemic information is usually essential to identify the word.

More interesting is the case of the word that is only fairly familiar. In this case it is likely that semantic associations serve only to narrow down the field, quite rapidly, no doubt, to a group of semantically related entries. At this point, a reader who cannot exploit the internal morphophonemic structure of the words has no alternative but to guess, and poorly trained or aphasic readers will often substitute a semantically related word for the correct one. But for the reader who can use this internal structure, lexical search is unambiguous and self-terminating; the word is there, with a morphophonemic representation consistent with its orthographic form, or it is not.

It has been shown experimentally that the internal structure of words facilitates recognition, and continues to do so even after the words have become quite familiar (Brooks, 1977). One might suppose that the more advanced the material being read, the more often the reader would be reading low-frequency, "fairly familiar" words, and the more important the ability to exploit the morphophonemic information in the orthography would become.

It would appear, then, that it would be to the advantage of a reader to be phonologically mature, to know the phonology of the language, so that the morphophonemic representations of words in his personal lexicon match the transcriptions of the orthography. If he is phonologically mature, he has, in the course of acquiring English, mastered the Laxing, Diphthongization, and Vowel Shift rules, and he has inferred that [hīyl] and [helθ] can both be derived from /hēl/, /θ/ being a separate morpheme. Thus he has /hēl/ and /hēl+θ/ as morphophonemic representations in his lexicon and not /hīyl/ and /helθ/. If he has not in fact gone through this process, the spellings heal and health will presumably seem to him arbitrary rather than regular.

This knowledge is, of course, a form of what has earlier been called "grammatical" knowledge, and it is of great significance that such knowledge is directly exploited in reading but not in listening. As has already been suggested, it is unlikely that the listener reconstructs the morphophonemic representations of words. As long as his lexical search strategy has paired the phonetic forms [hīyl] and [helθ] with their lexical entries, he can

analyze and understand the sentence. If his morphophonemic representations are immature, the only consequence is that the semantic information in the entry for health may not be as rich: He does not associate "health" with "healing."

Yet it would seem that for both beginning and experienced readers, access to morphophonemic representations is of even more importance than the maturity of these representations. The need for such access does not arise for the listener understanding sentences, presumably because he has innate automatic mechanisms for lexical search. If a reader has such access, that is, if he can bring his grammatical knowledge to bear on the task of reading, then the orthography will seem like a rational way of transcribing utterances in his language. Without access to grammatical knowledge, not only particular spellings, but the very idea of transcribing an utterance segmentally, will seem strange and arbitrary.

The state of having access to one's grammatical knowledge is what I meant by linguistic awareness[1] in my earlier paper (Mattingly, 1972). At that time I believed that this awareness had a metalinguistic, somewhat unnatural character. It now seems to me that such a state of awareness is eminently natural, since it is a mental state resembling that of the language-learner. The language-learner has access to morphophonemic representations because he is in the process of establishing them. Practical orthographies presuppose that the reader has the same sort of access to these representations.

To return to the question of what differences in the reading process are implied by differences in orthographic type, it would seem that what is primarily involved is the degree of linguistic awareness required. Logographic systems are the least demanding in this respect, since access only to morphological and not to the morphophonemic aspects of the representation is required, but the obvious price paid is that a large set of characters must be remembered. Alphabetic systems, on the other hand, are the most demanding. For language with appropriate morphological and phonological properties, a syllabary appears to be a happy compromise.

SOME EXPERIMENTAL EVIDENCE: PHONETIC RECODING

The orthographical evidence, then, suggests that a reader uses his grammatical knowledge to establish the lexical content of the sentence. But such evidence suggests nothing about how, given this information, the reader is able to understand the sentence, that is, how he reconstructs the phrase-marker and the semantic representation. He might conceivably make use of grammatical knowledge; he might use some analytic mechanism peculiar to reading and quite independent of spoken-language analysis; or he might use the analytic mechanisms that the listener uses.

The grammatical interpretation of earlier parts of a sentence depends generally on information in later parts. Since spoken sentences are physical events in real time, a listener must have a way of representing this early information in memory until he is prepared to analyze it. Yet to represent physical events in memory at all requires some analysis of these events. Thus the listener is compelled to make a rapid preliminary analysis that can then be deepened and refined in light of later information.

This preliminary representation is stored in short-term memory. Analysis of errors in short-term recall suggests that the information being stored is phonetic (Wickelgren, 1965a, 1966). However, since other sorts of linguistic information must obviously be stored in short-term memory as well (Fodor et al., 1974: Chap. VI), it would be more cautious to say that the short-term representation is at least phonetic.

Reading is a real-time process, just like listening. It is hardly relevant that the lexical information remains before the reader on the page. Whatever the analytic mechanisms he uses, he must make use of the results of earlier analysis in the course of current analysis, and it does not seem to occur to him to note his preliminary results in the margin. Thus, like the listener, he requires some form of temporary storage. Iconic storage, in which visual information is initially represented, is unsuitable for the purpose because of its very brief duration (Sperling, 1960). One might entertain the possibility of an "orthographic" short-term memory, analogous to "phonetic" short-term memory; or of a "semantic" short-term memory, in which words were represented by their meanings. Many individuals, however, report "inner speech" while reading, and some readers engage in actual articulatory or acoustic activity. These observations suggest that "phonetic" short-term memory itself provides temporary storage of information during reading.

There is, in fact, considerable evidence that if, in an experimental situation, orthographic material is to be temporarily remembered, "phonetic recoding" occurs (for a review, see Conrad, 1972). One experimental paradigm is considered in detail here because it provides opportunities for both semantic short-term memory and orthographic short-term memory to manifest themselves (Waugh & Norman, 1965; Kintsch & Buschke, 1969). In this paradigm a subject is asked to remember a list of words presented one by one, fairly rapidly. His recall of the list is then immediately tested by presenting a "probe" word and asking him to report the word that preceded the probe on the list. The typical finding is that words appearing early on the list and words appearing near the end of the list are better recalled than words in intermediate position. Recall of the later words is ascribed to a short-term memory representation still available at the time of the probe; in the case of intermediate words, this representation has decayed. Recall of the earlier words is ascribed to the subject's attempt to retain the list in long-term memory. When a list consists of semantically similar items, long-term but not short-term recall is reduced; when a list consists of phonetically similar items, short-term but not long-term recall is reduced. No effect is observed for words that are orthographically but not phonetically similar. These effects are the more impressive in that the phonetic and semantic similarities of the items are obvious to the subject, and he is free to use any available mnemonic strategy.

The effect of semantic similarity on long-term recall suggests that long-term memory is semantically structured, as one would expect. The effect of phonetic similarity on short-term recall suggests that the short-term memory representations are phonetically structured. This interpretation is corroborated by analysis of errors in other experiments on short-term recall of orthographic material. Such analysis reveals phonetic confusions similar to those observed in short-term recall of spoken material (Wickelgren, 1965b). The most reasonable inference is that the same short-term memory is used for

both types of material. The phonetic similarity effect, considered together with the absence of other short-term effects, also suggests that there is no equally convenient alternative to phonetic short-term memory with an appropriate duration--no orthographic or semantic short-term memory, for example. Had such alternatives been available, the subjects in Wickelgren's test could have avoided the disadvantages of storing phonetically similar items in phonetic short-term memory.

It is of considerable interest that the effects described are obtained not only for alphabetic but also for logographic stimuli. If the logographic kanji characters used in writing Japanese are presented to native speakers of Japanese in a probe paradigm, results parallel to those already described for English are obtained: When a list consists of kanji with phonetically similar readings, and only then, short-term recall is adversely affected (Erickson, Mattingly, & Turvey, 1972, 1977). Comparable results have been obtained for Chinese characters (Tzeng, Hung, & Wang, 1977).

To the extent that inference from the recall of visually-presented lists to actual reading is justified, these results suggest that orthographic information in reading is indeed "phonetically recoded" and stored in the short-term memory used for spoken language.

The validity of the inference to actual reading is strengthened by a related experiment. Liberman, I., Shankweiler, Liberman, A. M., Fowler, and Fischer (1977) tested the short-term recall of children considered good readers and children considered poor readers. Strings of five letters were briefly presented and after each presentation the subjects were asked to write down the letters in their given order. It was found that the performance of the good readers was better than that of the poor ones. More interestingly, it was found that the recall of a string in which the letter-names rhymed was poorer, relative to the recall of a control string, for good readers than for poor readers, presumably because the good readers more consistently employed phonetic short-term memory to retain the strings. Thus there seems to be a direct relationship between reading ability and phonetic recoding.

What is the significance of phonetic recoding? It is sometimes assumed--indeed, the term itself implies as much--that phonetic recoding takes place because of the supposed "phonetic" character of orthographies. According to this view, the reader converts letters (or letter patterns) to sounds, representing this information in short-term memory. This view is appealing because it seems to lead to a tidy statement of the relationship between reading and listening: In both activities, a phonetic representation is established that then serves as the basis for subsequent analysis.

But this view is unsatisfactory for several reasons. First, it has been shown that neither logographies nor alphabetic orthographies are phonetic transcriptions; yet phonetic recoding occurs with both. Second, this view ignores the lexical character of all practical orthographies. What they give the reader are word-identities, and hence, the phonological, syntactic and semantic information in lexical entries. Third, this view implies that in the analysis of a spoken sentence, the speech-perception mechanism is separate from, and does not interact with, other analytic mechanisms.

Phonetic recoding has no connection with the character of the orthography. As Kleiman (1974) has demonstrated, it does not occur simply in consequence of word recognition. What phonetic recoding means is rather that reader and listener are almost certainly employing the same analytic mechanisms. If the use of phonetic short-term memory reflects the reconstruction of the phonetic representation by these mechanisms in the course of analysis, it is very likely that they are doing the rest of the job as well.

Since the input to the analytic mechanisms in reading is lexical, and the required output is semantic, it might seem strange that the reconstruction and temporary storage of the phonetic representation cannot be dispensed with, as it is in artificial schemes for understanding printed text. But the analytic mechanisms constitute an intricate special-purpose system for understanding spoken sentences by working out their surface structure. The phonetic representation, though not logically required in reading, is an integral part of the product.

This point is clear, indeed, from the probe experiments. What the subject tries to do is to remember an entire list of words. The only way he can do this is to form semantic representations in long-term memory, in other words, to treat each item on the list as a one-word sentence to be analyzed, and he succeeds in doing so for the early words on the list. The formation of the short-term phonetic representation appears to be an essential part of the process. On the other hand, subjects who do not give evidence of phonetic recoding, like Liberman et al.'s (1977) poor readers, are probably not using any of their analytic mechanisms. If the formation of a phonetic representation could be readily dispensed with, Liberman et al. would have found <u>good</u> readers who gave no evidence of phonetic recoding.

It must be emphasized that the phonetic representation formed by a listener or a reader is abstract, like other mental representations, not a reenactment of speech production. Silent rehearsal, actual articulatory movement or subvocalization--forms of behavior that are with some justice regarded as marks of a slow and inefficient reader--are not an essential aspect of the phonetic representation. They are rather evidence that, for some reason, analysis of the sentence is proceeding so slowly that the information in short-term memory needs to be refreshed. Obviously, a skilled reader has no reason to employ such devices, but their absence does not suggest the absence of phonetic recoding. As for the subjective phenomenon of "inner speech," I do not know whether it is to be regarded as merely the consciousness of the phonetic representation or as a form of rehearsal. Since some very skilled readers report it, I am inclined to the former conclusion. But in any case, though it is no doubt a consequence of phonetic recoding, it is not a necessary consequence.

It is also often assumed, because of a similar misunderstanding, that the phenomenon of phonetic recoding means that reading speed is constrained by the relatively low rates at which speech can be uttered or the somewhat higher rates at which speech, if the signal is specially manipulated, may be understood. There is no justification for this view. How fast phonetic representations pass through short-term memory must depend on the rates at which orthographic forms can be recognized and sentences can be analyzed, not on the rates of speech production or speech perception.

READING AND LANGUAGE ACQUISITION: SOME CONJECTURES AND CONCLUSIONS

From the point of view adopted in this paper, the real mystery is that the analytic mechanisms can be used in reading at all. It has been stressed that these mechanisms are innate, highly specialized and inaccessible; it would seem entirely reasonable if it were the case that only a spoken sentence could be understood. Yet in reading, the boundary between grammatical knowledge and performance is crossed. The accessible information carried by the orthography is somehow able to trigger these inaccessible processes. Moreover, this information is not at all equivalent to the auditory information that initiates the process of understanding speech; as has been emphasized, it is lexical information.

My explanation, which must be regarded as purely conjectural, is that the reader takes advantage of a language-acquisition procedure. Part of the task of a language-learner, of course, is to increase his stock of lexical entries. Consider how he might go about this. Suppose that his data consist of the phonetic representation and the phrase-marker of a sentence containing a word that is new to him, and that in addition he has gathered from the situational context a tentative notion as to the semantic representation. To explain these data, he hypothesizes the existence of a word whose lexical entry has certain semantic, syntactic and phonological features. To test his hypothesis, he supplies the analytic mechanisms with this postulated lexical information, as well as with previously determined lexical information about other words in the sentence under consideration. If the analysis yields the observed phonetic representation and the assumed semantic representation, the proposed lexical entry is corroborated. That the analytic mechanisms can accept lexical information as input is perhaps not too surprising. It has already been argued that understanding speech is a nonlinear process with semantic and syntactic as well as auditory input.

This quite speculative but, I feel, not totally implausible account of what must be involved in learning a new word assumes that the language-learner is innately equipped to initiate the analysis of a sentence, starting with lexical information that is accessible to him. If we are willing to believe that the reader exploits this aspect of his language-learning capacity, a tentative explanation is available of the reader's otherwise surprising ability to make use of mechanisms that might be supposed to be reserved for the listener.

In an earlier section, it was observed that orthographies appealed to grammatical knowledge, that is, to knowledge of the language in the form that it is acquired by the speaker-hearer. If the reader is to recognize words efficiently, it is important for him not only to have such knowledge (phonological maturity) but to have access to it (linguistic awareness), just as he did when the knowledge was originally acquired. The present discussion has led us in the same direction. It has been suggested that the reader's ability to make use of the analytic mechanisms is part of his capacity for language acquisition. If this is really the case, then the resemblance between the reader and the language-learner can be extended. Not only do they both have access to grammatical knowledge, but also they both activate the analytical mechanisms with lexical information.

It has been mentioned already that the course of language acquisition varies considerably. At one extreme is the individual who virtually abandons language acquisition as soon as he has developed the relatively modest body of analytic strategies needed to get by; at the other extreme, the "word-child" who continues indefinitely to add to his grammatical knowledge.

I now offer the further conjecture that differences in children's readiness and ability to learn to read are related to these different patterns of language acquisition. The child who is still actively acquiring language at the time he begins to read will be relatively mature phonologically, so that the orthography will correspond to a considerable extent with his morphophonemic representations. Having access to these representations, he will be linguistically aware, and the orthography will seem to him a plausible way of representing sentences. The analysis of a sentence on the basis of its lexical content will present no problems, since he has continued to use this analytic procedure in the course of learning new words. Moreover, he will, as a linguist, see that reading is a source of fresh data. If he does not already have the morphophonemic forms /hēl/ and /hēl+θ/ in his lexicon, and the associated rules in his phonology, the orthographic forms heal and health will prompt him to revise his grammar accordingly. Thus the linguistic curiosity that has motivated his continuing language acquisition will motivate his learning to read as well.

On the other hand, the child who is no longer very actively acquiring language will surely find learning to read very difficult and unsatisfying. His morphophonemic representations will be less mature than they might be, so that the discrepancies between the orthography and the morphophonemic representations will be substantial. More seriously, these representations, being part of his grammatical knowledge, will have become less accessible to him; he will be lacking in linguistic awareness. As a result, the orthography will seem a mysterious and arbitrary way to represent sentences. Finally, since his capacity for language learning will not have been recently exercised, he may well have lost some of his ability to analyze a sentence on the basis of its lexical content.

Because most people continue to be capable of learning new words when they must, and even new languages when circumstances compel them to, it does not seem likely that the capacity for language acquisition atrophies completely. If not, there is certainly reason to hope that in poor readers this capacity is merely dormant, a muscle that has grown flabby from disuse. With proper instruction and appropriate environmental stimulation, it can be reawakened. But obviously, it would be even better if language acquisition had not been allowed to falter in the first place, if there had been no awkward interval between the period of learning to talk and the period of learning to read. This observation is not to be construed as a demand for very early reading instruction, but rather as a plea for linguistic stimulation above and beyond speaking and listening during pre-school years: story-telling, word-games, rhymes and riddles and the like. The value of such stimulation is certainly appreciated by most specialists in reading. But its justification is not merely that it prepares the child for the experience of learning to read, but that it helps to keep active psychological mechanisms that are indispensable in learning to read.

To summarize what must appear to be a rather discursive argument, it is my contention that written language, far more than speech, places a direct demand on the individual's acquired knowledge of language--what has been called grammatical knowledge; and that such knowledge, consequently, must be accessible to the reader, as it presumably is to the language-learner. Moreover, it appears that although reading and listening use the same analytic mechanisms--hence "phonetic recoding"--analysis of a sentence is accomplished in reading, unlike listening, from an input that corresponds to the lexical content of the sentence. It is conjectured that the reader is able to do this by means of what is really a language-acquisition procedure. Reading thus has much in common with language acquisition, and the child who has continued to acquire language beyond what is required for performance is more likely to learn to read easily than the child whose language acquisition capacity has become dormant.

REFERENCES

Bell, A. M. Visible speech: The science of universal alphabetics. London: Simkin Marshall, 1867.

Brooks, L. Visual patterns in fluent word identification. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading. Hillsdale, N.J.: Erlbaum, 1977.

Chomsky, N. Comments for Project Literacy meetings. Project Literacy Report No. 2, pp. 1-8, 1964. Reprinted in M. Lester (Ed.), Readings in applied transformational grammar. New York: Holt, Rinehart and Winston, 1970.

Chomsky, N. Aspects of the theory of syntax. The Hague: Mouton, 1965.

Chomsky, N. Reflections on language. Glasgow: Fontana/Collins, 1975.

Chomsky, N. Essays on form and interpretation. New York: North Holland, 1977.

Chomsky, N., & Halle, M. The sound pattern of English. New York: Harper and Row, 1968.

Conrad, R. Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Erickson, D., Mattingly, I. G., & Turvey, M. Phonetic coding of Kanji. Journal of the Acoustical Society of America, 1972, 52, 132.

Erickson, D., Mattingly, I. G., & Turvey, M. Phonetic activity in reading: An experiment with Kanji. Language and Speech, 1977, 20, 384-403.

Fodor, J. A., Bever, T. G., & Garrett, M. F. The psychology of language. New York: McGraw-Hill, 1974.

Harris, J. W. Spanish phonology. Cambridge, Mass.: MIT Press, 1969.

Kintsch, W., & Buschke, H. Homophones and synonyms in short-term memory. Journal of Experimental Psychology, 1969, 80, 403-407.

Kleiman, G. N. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1974, 14, 323-339.

Klima, E. S. How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Liberman, I., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. P. Orthography and the beginning reader. Paper presented at Cross-Language Conference on Orthography, Reading and Dyslexia, Bethesda, MD., September

18, 1978. Haskins Laboratories Status Report on Speech Research, 1979, SR-57, 67-84.

Liberman, I., Shankweiler, D. P., Liberman, A. M., Fowler, C., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), Towards a psychology of reading. Hillsdale, N.J.: Erlbaum, 1977.

Martin, S. Nonalphabetic writing systems: Some observations. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Mattingly, I. G. Reading, the linguistic process and linguistic awareness In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Mattingly, I. G. The psycholinguistic basis of linguistic awareness. Paper presented at Twenty-Eighth National Reading Conference, St. Petersburg, Fl., December 8, 1978. To appear in Conference Yearbook. Haskins Laboratories Status Report on Speech Research, 1979, SR-57, 155-160.

Moskowitz, B. A. On the status of vowel shift in English. In T. Moore (Ed.), Cognitive development and acquisition of language. New York: Academic Press, 1973.

Sperling, G. The information available in brief visual presentations. Psychological Monographs, 1960, 74 (No. 498).

Tzeng, O. J. L., Hung, D. L., & Wang, W. S.-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 621-630.

Waugh, N. C., & Norman, D. H. Primary memory. Psychological Review, 1965, 80, 1-52.

Wickelgren, W. A. Distinctive features and short-term memory for English vowels. Journal of the Acoustical Society of America, 1965, 38, 583-588. (a)

Wickelgren, W. A. Short-term memory for phonemically similar lists. American Journal of Psychology, 1965, 78, 567-574. (b)

Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustical Society of America, 1966, 39, 388-398.

Wilkins, J. An essay towards a real character and a philosophical language. London: S. Gellibrand, 1668.

FOOTNOTE

[1]My distinction between phonological maturity and linguistic awareness is perhaps slightly artificial. Klima (1972) has suggested that the morphophonemic representation may be less accessible than certain shallower levels of derivation. If so, it would be difficult to distinguish empirically between lack of phonological maturity and lack of linguistic awareness, especially in the case of "nonproductive" phonological rules, like Vowel Shift. (Moskowitz [1973], however, appears to have surmounted this difficulty.) Pending clarification, we are assuming, parsimoniously, that linguistic awareness means having access to the appropriate units of one's morphophonemic representations, while phonological maturity means controlling the phonological rules and having morphophonemic representations in one's lexicon approximating those of an ideal speaker-hearer of one's language.

A RANGE-FREQUENCY EFFECT ON PERCEPTION OF SILENCE IN SPEECH

Bruno H. Repp

Abstract. The amount of silence between VC and CV syllables (such as /ib/ and /ga/) was varied in a number of steps to yield percepts ranging from VCV to VCCV. These stimuli were presented in two conditions in which one or the other endpoint stimulus (the anchor) occurred on 30 percent of the trials, with a 30-percent restriction of the range of silence durations at the other extreme. Listeners' VCV/VCCV boundaries shifted dramatically between the two anchoring conditions. The unexpectedly strong context sensitivity of the silence cue has obvious methodological implications. It also suggests that perception of phonetically significant silence durations in speech is not limited by any psychophysical threshold but, rather, is guided by a specifically phonetic interpretation of the acoustic signal.

INTRODUCTION

Effects of stimulus environment on the perception of (phonetically ambiguous) speech sounds have been amply documented. Besides numerous studies of selective adaptation, there have been demonstrations of effects of stimulus range (e.g., Brady & Darwin, 1978; Rosen, 1979), stimulus frequency (e.g., Simon & Studdert-Kennedy, 1978), and sequential contrast (e.g., Diehl, Elman, & McCusker, 1978; Repp, Healy, & Crowder, 1979), among others. For a discussion of how these phenomena might be related, see Simon and Studdert-Kennedy (1978). While the existence of such context effects is no longer in doubt, their magnitude for different types of stimuli and perceptual cues is of theoretical and methodological interest.

In general, the above-mentioned effects are small in magnitude. However, this is frequently a consequence of the small region of perceptual uncertainty on a stimulus continuum, as well as, perhaps, of certain limits to the auditory perception of the relevant cue dimension. For example, the voiced-voiceless distinction on a synthetic voice-onset-time (VOT) continuum tends to be rather sharp, leaving little room for response shifts since, in general, only responses to stimuli close to the perceptual boundary are affected by context. Also, perception of VOT, as a brief noise-filled interval, may have a lower bound (i.e., a psychophysical detection threshold) that sets a limit

---

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories. Thanks are due to Patti Price for running Experiment II and analyzing the data of both experiments, and to Michael Studdert-Kennedy for helpful comments on an earlier draft of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

to downward shifts of the category boundary. The finding that vowels tend to exhibit larger context effects than consonants distinguished by VOT or by formant transitions (Eimas, 1963; Simon & Studdert-Kennedy, 1978) is in agreement with these ideas: Vowel continua also tend to exhibit larger regions of response uncertainty, and there are no obvious constraints on the auditory perception of the relevant cues. In this respect, vowels behave almost like nonspeech dimensions such as pitch or loudness that, too, are highly susceptible to contextual effects (Helson, 1964; Sawusch & Pisoni, 1974; Sawusch, Pisoni, & Cutting, 1974; Simon & Studdert-Kennedy, 1978).

The purpose of the present studies was to investigate the extent of combined range-frequency effects on the perception of silent intervals distinguishing single intervocalic stop consonants from sequences of two such consonants (Dorman, Raphael, & Liberman, 1979; Repp, 1978, 1979). This paradigm involves variations in the silent interval between syllables of the form VC_1 and C_2V; in the "single-cluster" distinction, $C_1 \neq C_2$, whereas, in the "single-geminate" distinction, $C_1 = C_2$. (The vowels may or may not be the same.) In each case, perception changes from VC_2V at short silent intervals to VC_1C_2V at longer intervals, but much longer periods of silence are required to achieve that distinction in the single-geminate case (ca. 200 msec) than in the single-cluster case (ca. 70 msec).

Although these distinctions have been investigated in a number of experiments (Dorman et al., 1979; Raphael & Dorman, in press; Repp, 1978, 1979), their degree of context sensitivity is not known. Repp (1979) showed that both perceptual boundaries vary considerably as a function of changes in properties of the stimuli themselves (such as spectrum, amplitude, and duration), but it is not known whether they are affected by the range and/or frequency of silence durations employed in a given stimulus ensemble. This issue is of more than methodological interest. If the perception of silence in these speech stimuli is limited by a psychophysical detection threshold, there should be a definite lower limit to perceptual boundary shifts as a function of context. Thus, if the single-cluster boundary of ca. 70 msec found in earlier studies had a psychoacoustic basis, it should not be susceptible to downward movement by, e.g., anchoring. (70 msec seems a bit high for a detection threshold, but the possibility cannot be ruled out *a priori*.) More generally, if the single-cluster boundary coincided with a psychoacoustic boundary of some sort, we might expect it to be relatively immune to contextual effects, simply because auditory limitations should not change a great deal as a function of stimulus environment. The single-geminate distinction was included here for comparison purposes; because of the much longer silence durations involved, psychoacoustic constraints were not likely to play a role. Essentially, therefore, the single-geminate distinction was expected to be highly context-sensitive; the question to be investigated was to what extent this is also true for the single-cluster distinction. A certain amount of sensitivity was to be expected from the fact that single-cluster boundaries tend not to be particularly sharp (Dorman et al., 1979; Repp, 1979).

A combined range-frequency paradigm was used, since range effects (due to shifting the entire range of stimulus values) and frequency (anchoring) effects (due to higher frequency of occurrence of one extreme stimulus) are likely to operate in similar ways. (In fact, both phenomena may derive

entirely from sequential contrast.) No attempt was made in the present studies to dissociate the relative contributions of these two effects which, in any case, depend on the precise ranges and anchor frequencies chosen.

EXPERIMENT I

Method

Subjects. Eight subjects participated. They included six paid student volunteers, one research assistant, and the author. All subjects had had some previous exposure to synthetic speech sounds.

Stimuli. Three syllables, /ib/, /ga/, and /ba/, were created on the OVE IIIc synthesizer at Haskins Laboratories. They had been previously used by Repp (1979) in related studies, and the reader is referred to that earlier paper for details of stimulus construction. In the single-cluster condition, /ib/ was followed by /ga/ after an interval of silence that varied between 15 and 115 msec in 10-msec steps. In the single-geminate condition, /ib/ was followed by /ba/ after an interval of silence that varied between 115 and 315 msec in 20-msec steps. Silent intervals were specified by computer instructions after digitizing the stimuli using the Haskins Laboratories Pulse Code Modulation (PCM) system.

Each of these two conditions was further subdivided into two range-frequency conditions. In the low-anchor condition, the stimulus with the shortest silent interval occurred 30 times, the stimuli with the three longest intervals did not occur at all, and the remaining seven stimuli occurred 10 times each. In the high-anchor condition, the stimulus with the longest silent interval occurred 30 times, the stimuli with the three shortest intervals did not occur at all, and the remaining stimuli occurred 10 times each. These two conditions are contrasted schematically in Figure 1. Note that the stimuli in the middle range (45-85 msec in the single-cluster condition, 175-255 msec in the single-geminate condition) were equally represented in each anchoring condition; their perception in the different contexts provided by the remaining stimuli provided the desired measure of context sensitivity.

The four stimulus sequences were recorded on magnetic tape. Each sequence included 100 test stimuli in random order, with interstimulus intervals (ISIs) of 2.5 sec. Each sequence was preceded by 10 examples: The two extreme stimuli (with silent gaps of 15 and 115 msec in the single-cluster condition, of 115 and 315 msec in the single-geminate condition) alternated five times.

Procedure. The tapes were played back on an Ampex AG-500 tape recorder at a comfortable intensity. The subjects listened in a quiet room over Telephonics TDH-39 earphones. The four tapes were presented in a single one-hour session. Their order was counterbalanced across subjects, with single-cluster and single-geminate conditions alternating. The subjects were instructed to write down "g" or "bg" in the single-cluster condition, "b" or "bb" in the single-geminate condition. The distinctions were illustrated by the sample stimuli at the beginning of each tape. The subjects were told not

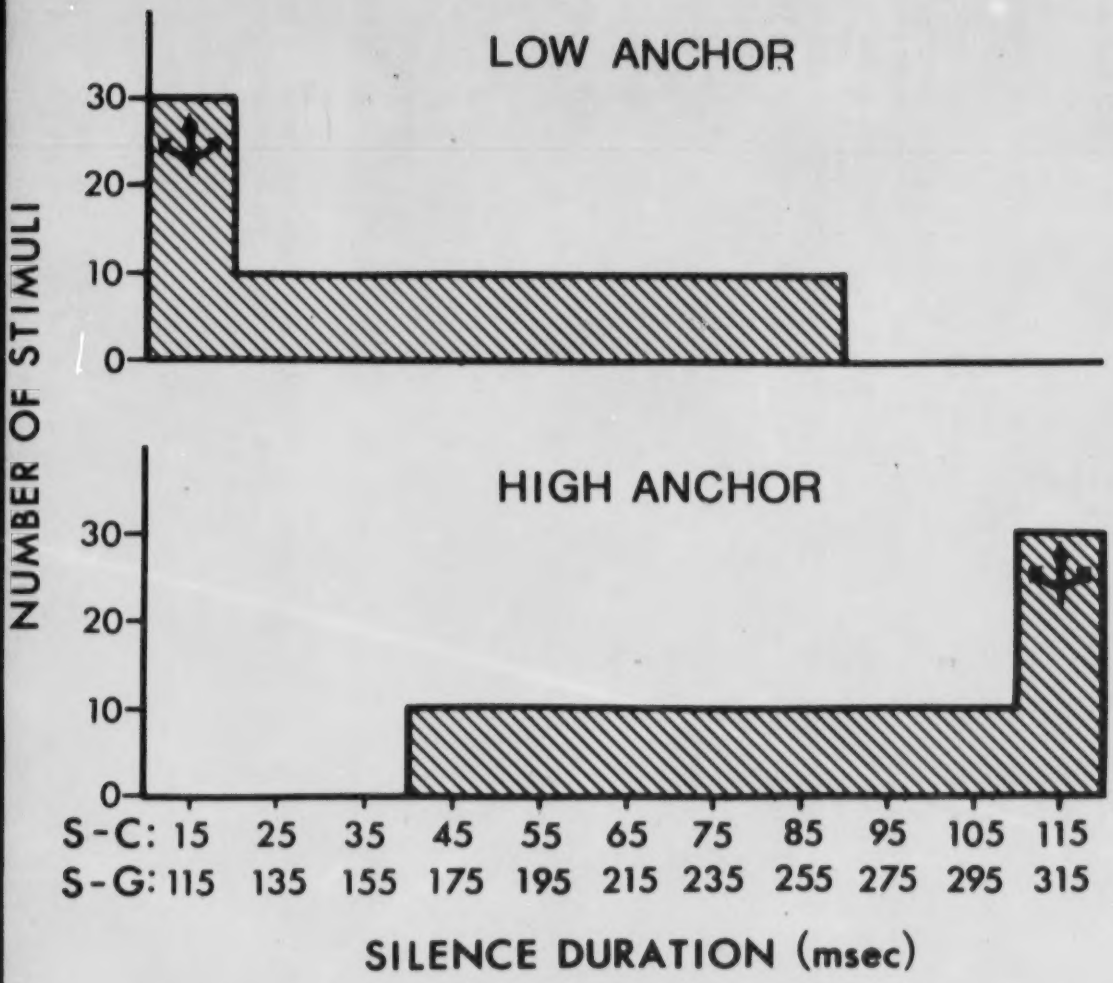


Figure 1. Schematic illustration of stimulus distributions in two range-frequency (anchoring) conditions.

to expect equal frequencies of the two response alternatives; however, they were not informed about the range-frequency manipulations. The author, of course, was an exception, and it was interesting to note that his foreknowledge by no means prevented large response shifts contingent on range and frequency.

Results and Discussion

The results are displayed in Figure 2 (solid lines). For comparison, the figure also shows data from earlier tests that included the full stimulus range, without anchoring (dotted line; from Repp, 1979, Exp. I). It can be seen that the range-frequency manipulations had a dramatic effect on subjects' responses. In the single-cluster condition, the 50-percent cross-over point of the average response function shifted from 44 msec in the low-anchor condition to 86 msec in the high-anchor condition--a difference of 42 msec. In the single-geminate condition (one subject's data were rejected here because of a nearly random response distribution), it shifted from 167 msec in the low-anchor condition to 233 msec in the high-anchor condition--a difference of 66 msec. All subjects tested showed these shifts, including the author who, despite foreknowledge and to his considerable surprise, was affected just as much as the other subjects. Comparison with the no-anchor data suggests that low and high anchors effected about equal response shifts up and down the scale of silence durations.

It is worth noting that, for three of the four anchoring conditions, the 50-percent cross-over point of the response function coincides almost exactly with the average silence duration of the stimuli in the sequence. These averages are 43 and 87 msec, respectively, for the single-cluster condition, and 171 and 259 msec, respectively, for the single-geminate condition. (Only the high-anchor single-geminate function is a little off.) Thus, it seems that, without explicit awareness, subjects somehow "computed" the weighted mean of the stimulus range and placed their category boundaries right there. This result, of course, is exactly what Helson's (1964) adaptation level theory predicts. A second prediction of psychophysical theory (Parducci, 1974) is also confirmed in Figure 2: The slopes of the anchoring functions (obtained with a restricted stimulus range) were steeper than the slope of the no-anchor function.

The finding of strong range-frequency effects in the single-geminate condition was not unexpected. The single-geminate distinction is not very familiar to English listeners (geminates occur only across morpheme boundaries and rarely are distinctive in English) and it involves a rather subjective judgment. Essentially, it may involve nothing more than judgments of the duration of the silent interval, even though experienced listeners (such as the author) feel they are making linguistic decisions. Thus, the perception of a single-geminate stimulus series might be expected to follow general psychophysical principles.

The really surprising result is the extent of the range-frequency effect on the single-cluster continuum. Here, the judgments are definitely linguistic, as they involve the distinction between one vs. two nonidentical phonetic segments. The task may also be conceptualized as requiring the detection of (the cues for) the first stop consonant, or the differentiation of these cues

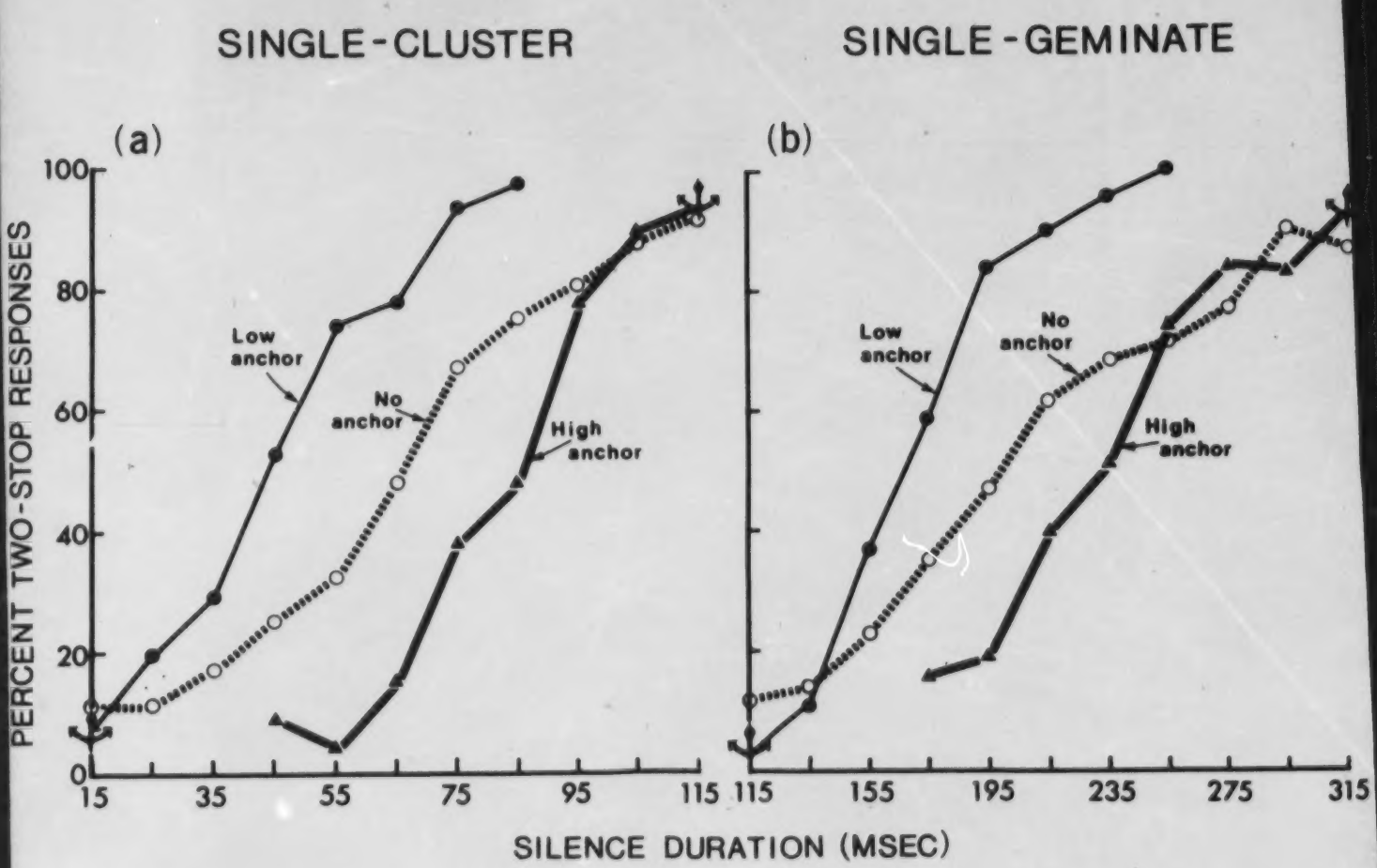


Figure 2. Range-frequency effects on the single-cluster and single-geminate distinctions. The no-anchor data are from Repp (1979, Exp. I).

from those for the second stop (Repp, 1979). The silence durations were rather short here, so that one might perhaps expect auditory interactions (such as backward masking) between the signal portions preceding and following the silent interval, as well as a psychophysical limit to the detectability of the silent interval. Both factors should constrain the extent to which the single-cluster boundary can be shifted to lower values of silence duration. However, if such constraints exist, it seems that they were not reached in the present study. Comparison with the no-anchor data (Figure 2) suggests that the boundary was shifted to lower values (using a low anchor) as easily as to higher values (using a high anchor). Thus, while certain psychoacoustic constraints may exist on the detectability of silence and of the cues preceding it, these constraints seem to play no role in the single-cluster distinction.

One obvious further point to consider is the possibility that the large range-frequency effects were simply a consequence of the large region of uncertainty, reflected in the shallow slopes of the identification functions. This question was investigated by computing product-moment correlations across subjects between slopes of individual identification functions and extent of boundary shift. While a negative correlation of -0.55 was found in the single-geminate condition (which supports the hypothesis that smaller slopes go with larger shifts), a positive correlation of +0.59 was found in the single-cluster condition (which contradicts the hypothesis); both correlations were nonsignificant. It should also be noted that individual identification functions were often considerably steeper than the average functions shown in Figure 2, and there were several instances of large boundary shifts despite steep slopes. Thus, no convincing evidence for a direct relation between uncertainty and sensitivity to range-frequency was found within the present experiment, suggesting that a shallow-sloped identification function is neither necessary nor sufficient for large context effects to occur. This does not rule out the possibility that there is a tighter relationship between these variables across different stimulus types.

EXPERIMENT II

The question now arises whether the observed response shifts reflect a true perceptual effect, or whether they were simply due to a common response bias--a tendency to give equal numbers of responses in each category (even though this would have been contrary to instructions). One indication that more is involved than a simple response bias is the fact that the author, an informed and experienced subject, showed the same shifts as the other, uninformed and less experienced listeners. To go beyond this preliminary evidence, Experiment II attempted to demonstrate range-frequency effects in a two-alternative forced-choice discrimination task. Only the single-cluster distinction was amenable to this approach: The first stop consonant, which is sometimes heard and sometimes is not, was varied to be either /b/ or /d/ (preceding /g/). Would the listeners' ability to discriminate these two alternatives at various silence durations be affected by range-frequency manipulations? On the basis of earlier data (Dorman et al., 1979; Repp, 1978), discrimination performance was expected to be poor (but perhaps better than chance) at short silence durations, and to improve rapidly with increasing silence duration. If range-frequency effects are truly perceptual (at

least in part), discrimination performance should be better in a low-anchor condition than in a high-anchor condition. This is so because a given silence duration should be effectively longer in the low-anchor condition, thus making the spectral cues for the first stop (/b/ vs. /d/) easier to detect and, hence, easier to discriminate, even if they do not lead to a separate phonetic percept.

A second purpose of Experiment II was to examine the generality of the findings of Experiment I by replicating them with natural-speech stimuli of somewhat different structure. Perhaps, it was the synthetic quality of the stimuli in Experiment I that produced such large range-frequency effects.

Method

Subjects. Eight subjects participated, including six paid student volunteers, one research assistant, and the author. Only the author had previously been a subject in Experiment I.

Stimuli. The stimuli were selected from a larger set of natural utterances that had been recorded for use in a different experiment. Included in this set were /ab-gu/ and /ad-gu/, recorded four times by a male speaker (the author). Thus, four different tokens of each utterance were available. The utterances were digitized and edited to eliminate the closure period. Varying periods of silence were then inserted in place of the original closure; durations ranged from 0-100 msec in 10-msec steps.

Low-anchor and high-anchor sequences were constructed that were similar to those in Experiment I. In the low-anchor condition, the 0-msec stimulus occurred three times as often as each of the other stimuli (with silences ranging from 10 to 70 msec), whereas, in the high-anchor condition, the 100-msec stimulus occurred three times as often as the other stimuli (with silences ranging from 30 to 90 msec). This was true for each of the four tokens of /ab-gu/ and /ad-gu/, resulting in a basic set of 80 stimuli ([3 repetitions of the anchor plus 7 other stimuli, each occurring once = 10] x 4 tokens x 2 utterances) which was recorded three times in different randomizations, with ISIs of 2.5 sec. A sample sequence preceded each anchoring sequence; as in Experiment I, these examples consisted of the extreme stimuli (0 and 100 msec of silence, respectively) in alternation. Eight such alternating pairs were provided, one for each token of each utterance.

Procedure. Each subject participated in two tasks involving the same tapes. The first task was three-alternative forced choice (identification); the response choices were "bg", "dg", and "g". The second task was two-alternative forced choice (discrimination); the response choices were "b" and "d", referring to the first consonant, whether or not a phonetic percept was available. The sequence of the two tasks was fixed for all subjects. The order of the low-anchor and high-anchor conditions was counterbalanced across subjects but constant within tasks for a given subject. The structure of the stimuli and the perceptual consequences of varying the silent interval were explained in detail to the subjects; thus, they were aware that there were always cues to a "b" or "d" at the end of the first stimulus portion, although these cues might not always lead to a phonetic percept. On the other hand, the subjects--with the exception of the research assistant and the author--

were not informed about the range-frequency variations.

Results and Discussion

The results of the three-alternative forced-choice task are plotted in Figure 3a as percentage of correct responses, averaged over tokens and subjects. There was a clear and systematic effect of range-frequency, in accordance with Experiment I. The effect was somewhat smaller here (a boundary shift of about 28 msec vs. 42 msec in Exp. I), perhaps due to the greater naturalness of the stimuli. Clearly, however, the results demonstrate that strong range-frequency effects are not limited to synthetic materials.

The results for the /ab-gu/ and /ad-gu/ series were extremely similar (Figure 3a). There was some systematic token variability which, however, need not concern us here since all individual tokens showed the range-frequency effect. The 50-percent cross-overs of the low-anchor functions (around 30 msec) fell in the vicinity of the mean silent interval of the series (28 msec); the high-anchor functions, however, crossed the 50-percent line below that mean value (72 msec), at 60 msec. It is interesting to note that the cross-over points were at silences 10-25 msec shorter than in Experiment I. This may itself have been a range effect due to the 15-msec downward shift of the total stimulus range (0-100 msec in Exp. II vs. 15-115 msec in Exp. I); however, it may also reflect stimulus-specific factors.

Figure 3a does not display errors, i.e., two-stop responses in which the first stop was misidentified. Misidentifications of /b/ as "d" were much more frequent than of /d/ as "b", and while the frequency of the latter did not vary systematically as a function of silence duration, the former were more frequent at short silence durations, as one might expect. This is shown indirectly in Figure 3b which plots the percentage of two-stop responses to /ab-gu/ that were correct responses [i.e., 100 x %"bg"/(%"bg" + %"dg")]. Interestingly, these conditional percentages reveal a range-frequency effect, too: For the same silence durations, correct responses to /ab-gu/ were more likely in the low-anchor condition than in the high-anchor condition. Since these are conditional percentages, the effect cannot be due to the difference in probability of two-stop responses between the two anchoring conditions. Rather, it indicates that discrimination of /b/ and /d/ (or, alternatively, the correct identification of /b/) was facilitated in the low-anchor condition. This anticipates the results of the second task.

The results of the two-alternative forced-choice task are shown in Figure 4a. The response pattern for /ab-gu/ stimuli confirms the results shown in Figure 3b: Correct responses increased as silence duration increased, but they increased more rapidly in the low-anchor than in the high-anchor condition. The results for /ad-gu/ stimuli were different: Accuracy was higher, especially at the shortest silence durations, and there was no range-frequency effect (perhaps even a small reversed effect).

There are two ways of looking at these data. One way is to consider /ab-gu/ and /ad-gu/ stimuli separately. It must then be concluded that the /d/ (or, rather, some characteristic auditory stimulus property associated with it) was easier to detect than the /b/--this confirms earlier results obtained with synthetic speech (Repp, 1978)--and that, possibly for that reason, only

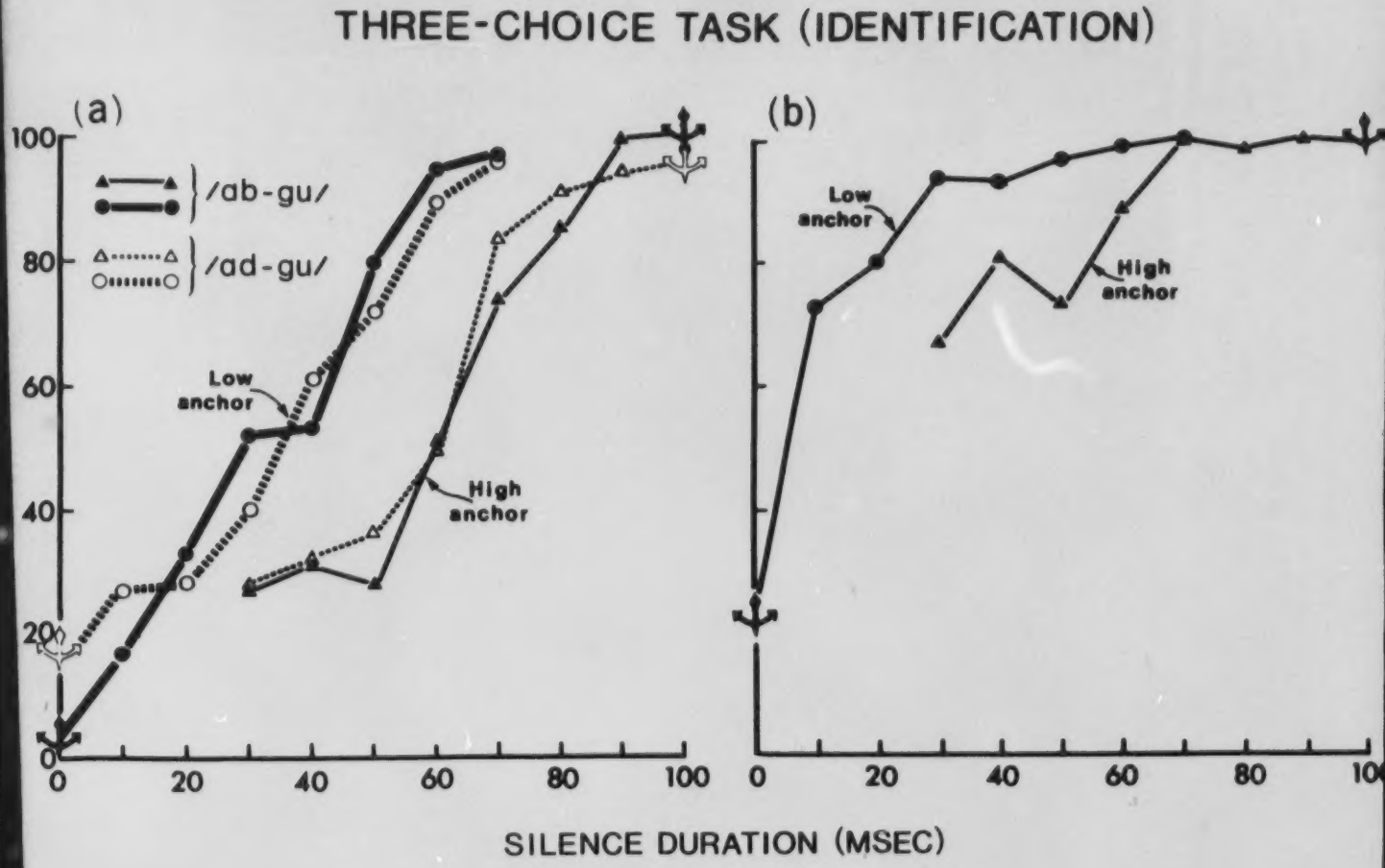


Figure 3. Range-frequency effects on the single-cluster distinction in three-alternative forced-choice identification. (a) Low- and high-anchor results for /ab-gu/ and /ad-gu/. (b) Percent correct of all two-stop responses to /ab-gu/, i.e., 100 x [%"bg"/%("bg" + "bd")].

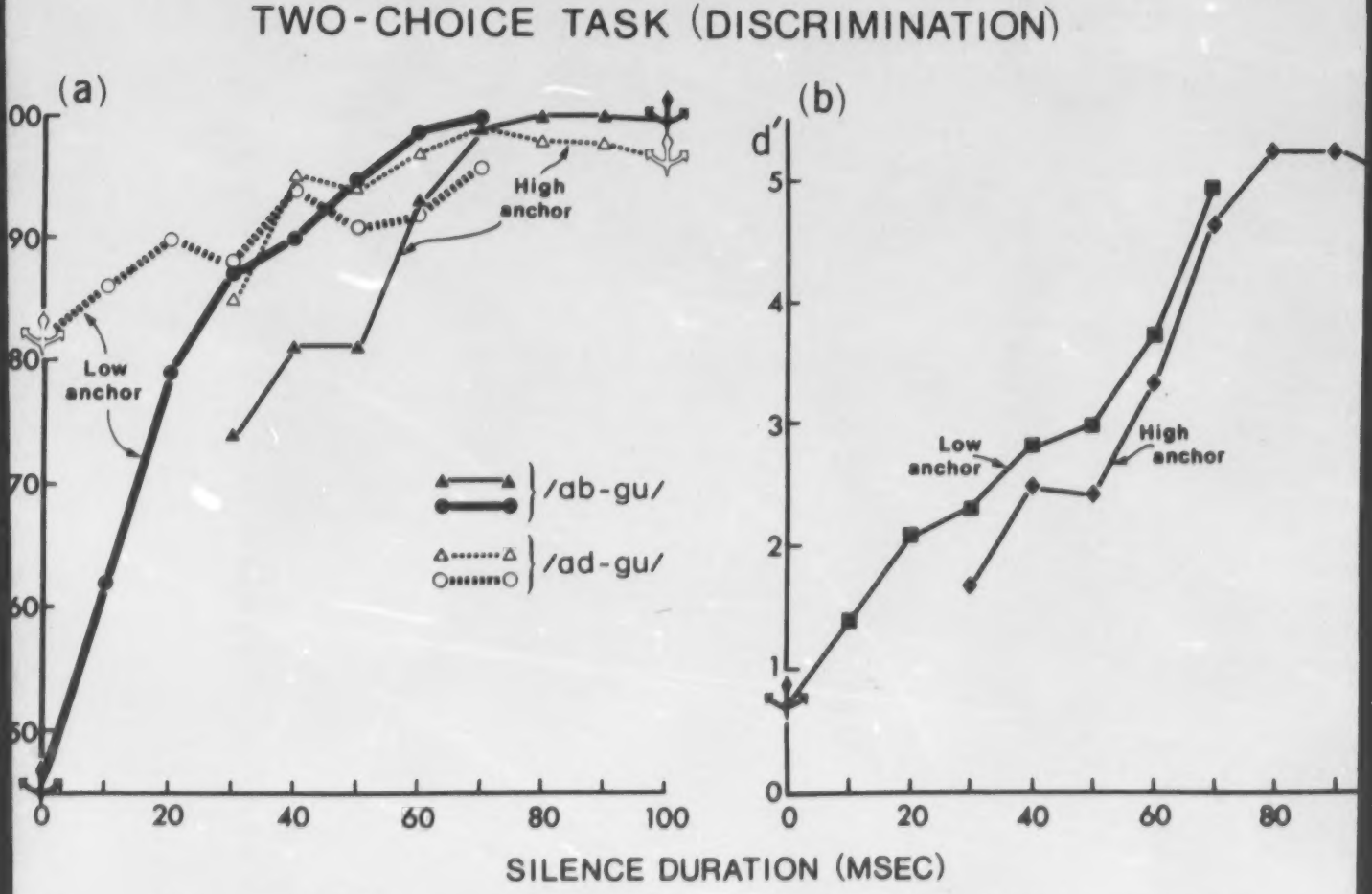


Figure 4. Range-frequency effects on the single-cluster distinction in two-alternative forced-choice discrimination. (a) Low- and high-anchor results for /ab-gu/ and /ad-gu/; chance = 50 percent. (b) The same results, expressed in terms of d' (d'_{max} = 6.40).

the detectability of /b/, but not of /d/, was affected by range-frequency manipulations. Alternatively, one might regard the apparent higher detectability of /d/ as being due to a bias in favor of "d" responses; this bias may have changed somewhat as range and frequency were varied, thus accounting for the absence of a range-frequency effect on /d/.[1] In this case, it is appropriate to combine the results of /ab-gu/ and /ad-gu/ to yield a single, bias-free measure of discrimination, such as d'. Values of d' were calculated from the average data, treating proportions of 0 and 1 as 0.001 and 0.999, respectively. (Thus, $d'_{max} = 6.40$.) These values are plotted in Figure 4b. There appeared to be a systematic effect of range-frequency condition on discriminability. Unfortunately, however, it did not reach significance. (Only four of eight subjects showed the effect.) Thus, the data merely suggest, but do not establish, that part of the range-frequency effect is perceptual in nature. Such a conclusion, if confirmed by further data, would conform to Helson's classical notion of adaptation level, shifts in which were taken to imply a change in the "sensory character" of the stimuli, in addition to shifts in judgment (Helson, 1964, p. 136). But, of course, the results are also in agreement with the well-known pattern of categorical perception: The availability of phonetic categories facilitates discrimination. Since a phonetic distinction emerged earlier in the low-anchor condition, discriminability was improved over the high-anchor condition (at least in some subjects who perhaps found it difficult to focus on the auditory properties of the stimuli).

GENERAL DISCUSSION

The extreme sensitivity of the single-cluster and single-geminate distinctions to stimulus range and frequency has obvious methodological implications. On the one hand, great caution will have to be exercised when using stimuli of this type in an experiment, especially when comparing results from different experiments--because category boundaries are likely to be unstable and sensitive to a variety of contextual factors. On the other hand, it is precisely this sensitivity that makes the single-cluster and single-geminate distinctions convenient for the investigation of various contextual effects, e.g., speaking rate (cf. Pickett & Decker, 1960; Repp, 1979). Therefore, the methodological consequences of the present results are not necessarily negative.

On the theoretical side, we have noted the absence of any obvious lower limit to the amount of silence required to perceive the first stop consonant in the single-cluster paradigm. Such a limit may exist at very short silences (0-20 msec), but the present experiments suggest that, with a more extreme range-frequency shift, the perceptual boundary might even be pushed into that region. There is evidence from related experiments that the acoustic cues specifying the first stop (viz., the formant transitions preceding the closure interval) "get through" in auditory processing even when no silence is present at all (Dorman et al., 1979; Ganong, 1975). Thus, it seems that these cues are not lost due to low-level auditory interference or integration. Rather, they seem to be available to a specifically phonetic decision process that chooses to interpret them as specifying an independent phonetic segment only when the signal parameters permit the inference that such a segment was actually intended by the speaker. It may be predicted, then, that certain conditions exist under which listeners perceive two stop consonants even in

the absence of any silence. At least one such condition has indeed been demonstrated. It involves a perceived change in speaker at the point where silence would normally occur; i.e., when VC is spoken by one voice and CV by another, listeners generally perceive both stops (Dorman et al., 1979). Another, more intriguing, condition is suggested by preliminary findings of Dechovitz (1979): It may be that, in sentence context, when a word boundary occurs with a major syntactic boundary between a VC-CV sequence, the silence is rendered perceptually ineffective. Such a top-down effect on phonetic perception was obtained by Dechovitz for silence as a cue to the fricative-affricate contrast; it remains to be demonstrated in the present paradigm.

It is also worth noting that, in another (unpublished) experiment I conducted, using stimuli nearly identical to the present set, a number of listeners required no silence to perceive both stop consonants. These individuals tended to be the less experienced listeners; experienced listeners such as the author himself were consistently unable to hear two stops at short silence durations. This puzzling result suggests that some listeners did not employ their phonetic-articulatory knowledge to its full extent; viz., they were not sensitive to the fact that two stops cannot be articulated in sequence without a sufficient closure of the vocal tract. However, the finding does support the hypothesis that auditory perception of the transitional cues and/or of the silent interval is not limited: It seems unlikely that basic auditory processes vary widely from individual to individual. Differences in perceptual strategy are indicated.

These observations suggest that the perceptual role of silence in speech is governed less by psychoacoustic principles than by an interpretative process sensitive to contextual factors (and perhaps subject to linguistic top-down constraints). Surely, psychoacoustic factors play a role to the extent that they determine the effective amount of silence that emerges from auditory processing (cf. Repp, 1979). However, the principles by which the auditory transform of the speech signal is decoded into phonetic segments are likely to be speech-specific. Range effects may arise at either the auditory or the phonetic level: they may reflect changes in the effective silence duration before any phonetic interpretation occurs, due to fatigue of certain auditory mechanisms; or they may reflect changes in the criterion that regulates acceptance to one or another phonetic category. The results of Experiment II contain a hint that more than a pure criterion shift is involved; however, a change in criterion probably accounts for the larger part of the range-frequency effect. This criterion change, in turn, could be contingent on the range or frequency of the relevant cue (silence duration) alone, or it could represent a more general perceptual normalization contingent on the perceived average speaking rate of the context. That is, the effects observed here may or may not be related to the boundary shifts observed when test stimuli are embedded in carrier sentences spoken at different rates of articulation (see, e.g., Repp et al., 1978). This issue could be investigated by producing changes in the perceived rate of the context by manipulating temporal cues other than silence itself.

Another question for future research is whether similarly large range-frequency effects are observed when silence acts as a cue for stop voicing (cf. Lisker, 1957, 1978), stop manner (cf. Fitch, Halwes, Erickson, & Liberman, in press), or for the fricative-affricate contrast (cf. Repp et al.,

1978). In the case of stop manner (e.g., slit-split) or the fricative-affricate distinction (e.g., say shop-say chop), the critical silence durations are typically rather short (20-40 msec) and labeling functions are steep, so that less context sensitivity might be expected. This seems to be true as far as speaking rate effects are concerned: Marcus (1978) has reported that the slit-split contrast is insensitive to variations in speaking rate, and Repp et al. (1978) reported rather small effects of speaking rate on the say shop-say chop distinction. Silence as a voicing cue (e.g., rabid-rapid), however, generally leads to boundaries in the same range as the /VgV/-/VbgV/ contrast; and, indeed, Port (1979) has reported sizeable effects of speaking rate on the voicing contrast. However, the precise relationship between the size of range-frequency effects on the one hand, and type of cue, function of cue, boundary location, and slope of the labeling function requires further study.

REFERENCES

Brady, S. A., & Darwin, C. J. Range effect in the perception of voicing. Journal of the Acoustical Society of America, 1978, 63, 1556-1558.

Dechovitz, D. R. Effects of syntax on the perceptual integration of segmental features. In J. J. Wolf & D. H. Klatt (Eds.), Speech communication papers presented at the 97th Meeting of the Acoustical Society of America. New York: Acoustical Society of America, 1979.

Diehl, R. L., Elman, J. L., & McCusker, S. B. Contrast effects on stop consonant identification. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 599-609.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 1979, 65, 1518-1532.

Eimas, P. D. The relation between identification and discrimination along speech and nonspeech continua. Language and Speech, 1963, 6, 206-217.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop-consonant manner. Perception and Psychophysics, in press.

Ganong, W. F., III. An experiment on "phonetic adaptation". Progress Report No. 116, 1975, 206-210. (M.I.T.: Research Laboratory of Electronics.)

Helson, H. Adaptation-level theory: An experimental and systematic approach to behavior. New York: Harper & Row, 1964.

Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. Language, 1957, 33, 42-49.

Lisker, L. Closure hiatus: Cue to voicing, manner, and place of consonant occlusion. Haskins Laboratories Status Report on Speech Research, 1978, SR-53 (vol. 1), 79-86.

Marcus, S. M. Distinguishing "slit" and "split"--an invariant timing cue in speech perception. Perception and Psychophysics, 1978, 23, 58-60.

Parducci, A. Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of Perception, Vol. II. New York: Academic Press, 1974. Pp. 127-141.

Pickett, J. M., & Decker, L. R. Time factors in perception of a double consonant. Language and Speech, 1960, 3, 11-17.

Port, R. F. The influence of tempo on stop closure duration as a cue for voicing and place. Journal of Phonetics, 1979, 7, 45-56.

Raphael, L. J., & Dorman, M. F. Silence as a cue to the perception of syllable-initial and syllable-final stop consonants. Journal of Phonetics, in press.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. Perception and Psychophysics, 1978, 24, 471-485.

Repp, B. H. Influence of vocalic environment on perception of silence in speech. Haskins Laboratories Status Report on Speech Research, 1979, SR-57, 267-290.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 129-145.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 621-637.

Rosen, S. M. Range and frequency effects in consonant categorization. Journal of Phonetics, 1979, 7, 393-402.

Sawusch, J. R., & Pisoni, D. B. Category boundaries for speech and nonspeech sounds. In Research on Speech Perception (Indiana University: Dept. of Psychology), Progress Report No. 1, 1974, 140-148.

Sawusch, J. R., Pisoni, D. B., & Cutting, J. E. Category boundaries for linguistic and nonlinguistic dimensions. In Research on Speech Perception (Indiana University: Dept. of Psychology), Progress Report No. 1, 1974, 162-173.

Simon, H. J., & Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. Journal of the Acoustical Society of America, 1978, 64, 1338-1357.

FOOTNOTE

[1]That this is indeed an alternative way of looking at the data, when dealing with linguistic stimuli, is illustrated by the following hypothetical case: Assume that /d/ is perfectly identifiable (at long silence durations, say) but /b/ still tends to be misidentified as "d" (e.g., because it is not a very good token); then, d' will always be maximal because of the 100-percent "hit" rate for /d/. Obviously, in this case the poor identifiability of /b/ cannot be ascribed to a response bias in favor of "d"; therefore, d' is not an appropriate measure. It is not clear whether the present situation (at short silence durations) fits that description since /b/ could be identified perfectly at longer silence durations. The unanswered question is: Did short silences make the /b/ sound like /d/ (i.e., did they act as a cue for an alveolar place of articulation, not unlike the way in which short intervocalic silences may cue a flapped /d/--cf. Lisker, 1978; Port, 1979), or did they create a general bias towards responding "d", which was unrelated to the acoustic cues in the signal?

PERCEPTION OF STOP CONSONANTS BEFORE LOW UNROUNDED VOWELS*

Ignatius Mattingly+ and Andrea Levitt++

Abstract. Previous experiments in the perception of stop-vowel syllables have sampled the entire vowel space rather coarsely. The present experiments look more closely at the perception of stops with four low unrounded vowels differing only in F2 frequency and heard as more or less backed variants of [a]. For each vowel, two labelling tests were prepared from synthesized stimuli. The onset of the F3 transition was varied in seven 200 Hz steps centering on previously obtained estimates of the [b-d] and [d-g] crossover points for F2 with a straight F3 transition. The pattern of crossover values obtained reflects the interaction of the F2 and F3 transition cues and the sharp differences in the velar locus before front and back variants of [a].

It is well known that perception of place in stops depends upon the transitions of the second and third formants (Liberman, Delattre, Cooper, & Gerstman, 1954; Delattre, Liberman, & Cooper, 1955; Harris, Hoffman, Liberman, Delattre, & Cooper, 1958; and Hoffman, 1958). Transitions for alveolar stops can be described in terms of fixed acoustic loci from which they may be regarded as originating, but the loci of the transitions for labial and velar stops vary depending on the adjacent vowel. For a labial stop the loci are lower than the vowel steady-state frequencies and increase as the steady-state frequencies increase. For velars, the F2 locus is high for low unrounded vowels and low for back rounded vowels. Moreover, these two cues are interdependent. Boundaries of perceptual categories for a set of F2 transitions to a particular vowel will shift if the F3 transition changes; conversely, boundaries for F3 categories will shift if the F2 transition changes.

The purpose of the experiments we conducted was to study this interaction of the steady-state frequency of F2 and the onset frequency for F2 and F3 transitions in the perception of the [b-d] and [d-g] boundaries for a narrow

---

*An earlier version of this paper was presented at the Ninth International Congress of Phonetic Sciences, 6-11 August 1979, Copenhagen, Denmark.

+Also, University of Connecticut, Storrs, Connecticut.

++Also, Wellesley College, Wellesley, Massachusetts.

Acknowledgment: Support from the National Institutes of Health (Grant HD-01994) and the Veterans Administration [Research Contract V101(134) P-342] is gratefully acknowledged.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

range of vowels. We found a systematic relationship between the F2 steady-state and the F2 and F3 transition onsets for the [b-d] boundary and a somewhat more complex pattern for the [d-g] boundary, a pattern which reflects the sharp difference in the velar locus before front and back variants of [a].

Previous experiments in the perception of stop-vowel syllables have sampled the entire vowel space rather coarsely (for example, Delattre et al., 1955). We wished to investigate the effect of simultaneously varying F2 and F3 transition onset over a narrower range of vowels. Three series of synthetic-vowel sounds were synthesized with the OVE III synthesizer. All three series had an F3 steady state at 2700 Hz. Each series had four members. Series I had an F1 steady state at 700 Hz and F2 values of 1000, 1160, 1400 or 1750 Hz. Series II had an F1 steady state at 770 Hz and F2 values of 1100, 1280, 1540, and 1925 Hz. Series III had an F1 steady state of 840 Hz and F2 values of 1200, 1400, 1680, and 2100 Hz. The F2 values in each series were generated on an approximately logarithmic scale.

Three tapes were prepared for labelling tests. Six phonetically trained subjects listened to the four tapes, each of which contained a randomized presentation of 20 repetitions of each of the four stimuli. For the 6 listeners, 71 percent of the stimuli of Series I were identified as low unrounded vowels; for series II, 84.6 percent of the stimuli were identified as low unrounded vowels, and for Series III, 91 percent of the stimuli were identified as low unrounded vowels. This last series gave the narrowest range of response variation and essentially allowed us to eliminate rounding as a factor. It was chosen for further investigation.

A number of experiments were conducted using this narrow range of vowels. Four stop-vowel continua were created by imposing a series of initial transitions 40 msec in length on each of the four vowels in the series. For F1, the frequency at onset was always 200 Hz; for F2, the onset frequency was varied from 850 to 2650 Hz in 10 steps, and for F3, the onset frequency was 2700 Hz, equal to the steady-state frequency. These stimuli were presented to nine subjects who were asked to label the stimuli as beginning with [b],[d], or [g]. After testing, the F2 [b-d] crossover estimates were 1250 Hz, 1350 Hz, 1650 Hz, and 1650 Hz, respectively, for the four vowels in the series. However, it seemed desirable to make more refined estimates of the [d-g] boundaries, which seemed more variable, so a further experiment was conducted with six subjects, in which the F2 onset was varied in nine 100 Hz steps centering on the [d-g] boundary estimates corresponding to the stated F2 steady-state frequencies. The revised crossover boundaries corresponding to the stated F2 steady-state frequencies were 2300 Hz, 2200 Hz, 2150 Hz, and 2200 Hz respectively.

After thus determining the [b-d] and [d-g] crossover points with only the F2 transition varying for the narrow range of vowels under consideration, we decided to see what shifts in the boundaries would occur with both the F2 and F3 formant onset frequencies changing.

For each of the four vowels then, two new sets of stimuli were prepared in which the F2 transition onset was varied in five 100 Hz steps centering on each of the two crossover estimates, the [b-d] and the [d-g], while the F3 transition onset was varied in seven 200 Hz steps centering on the F3 steady-

state frequency of 2700 Hz. There were thus eight sets of 35 distinct stimuli, one for the [b-d] crossover and one for the [d-g] crossover for each of the four vowels. Figure 1 is a schematic representation of two such sets of 35 stimuli, one set constructed around the [b-d] boundary and one set around the [d-g] boundary for a vowel with an F2 steady-state at 1680 Hz, the third vowel in the series we investigated. For each set of 35 test stimuli, a labelling test was prepared in which the 35 stimuli occurred eight times in random order. These tests were presented in a Latin square design to a group of twelve subjects. The subjects were instructed to label the stimuli as [b], [d] or [g].

Crossover values obtained for each of the four vowels from the resulting data are plotted as four graphs for /b-d/ and four graphs for /d-g/ in Figures 1 and 2. In each plot, the ordinate is the F3 transition onset and the abscissa is the F2 onset. Each point plotted represents the F3 crossover for a particular F2 transition onset. Note that in several cases, a particular F2 onset yielded the same response for all F3 onsets. For these cases no crossover point could be plotted.

The [b-d] plot in Figure 2 shows a systematic relationship among the three variables. In general, lower onset frequencies (i.e., more negative transitions) for either F2 or F3 yield [b] responses rather than [d] responses. The boundary value for either of the two onsets increases as the boundary value for the other decreases, so that the more negative the F3 transition is, the less negative the F2 transition need be to yield a [b] response. Increasing the steady-state frequency of F2 displaces the boundary toward higher F2 and F3 onset frequencies.

The [d-g] data in Figure 3 are somewhat more complex. In general, higher onset frequencies for F2 and lower onset frequencies for F3 yield [g] responses rather than [d] responses. The boundary values for the two onsets increase together, so that the more negative the F3 transition is, the less positive the F2 transitions need to be to produce a [g] response. With an increase in the F2 steady-state frequency, the boundary value is at first displaced toward lower F2 onset frequencies and lower F3 onset frequencies. But as indicated by the several crossings of the boundary value graphs for the lower three F2 steady-state frequencies, the displacement is not entirely consistent, and in the case of the highest two F2 steady states, the direction of the displacement is reversed for F2: The F2 boundary values are higher and the F3 boundary values are lower for the higher F2 steady state. This reversal reflects a sharp difference in the F2 loci for velars before front versus back variants of [a].

These results are quite consistent with the results of earlier studies cited above of the F2 and F3 transitions as cues to place, and with the acoustic theory of speech production as decribed by Fant (1960), but they explore in somewhat more detail the part of the vowel space associated with the velar locus difference. They confirm that a lower locus for [g] is associated with backness, as well as roundness in the following vowel.

The results of this experiment also demonstrate a trading relationship in that, within certain limits, changes in one cue can be compensated for by changes in another cue. The data from this experiment indicate, in fact, that

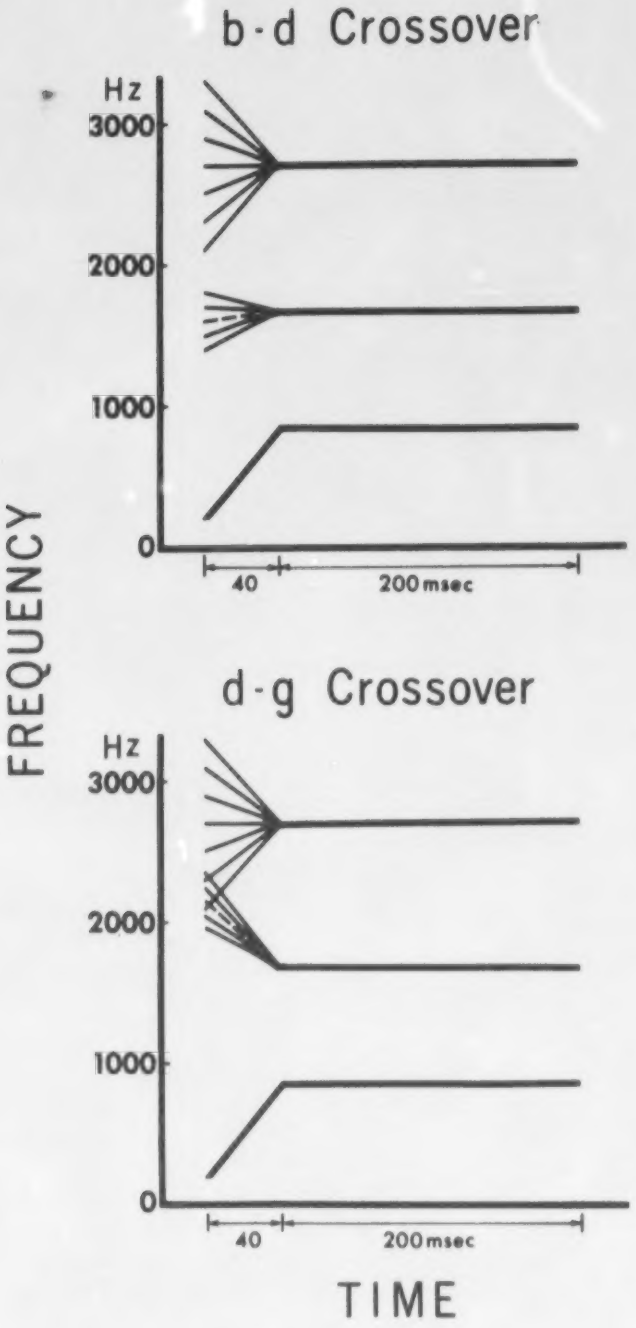

b-d Crossover
Hz
3000
2000
1000
0
40
200 msec
FREQUENCY
d-g Crossover
Hz
3000
2000
1000
0
40
200 msec
TIME

Figure 1

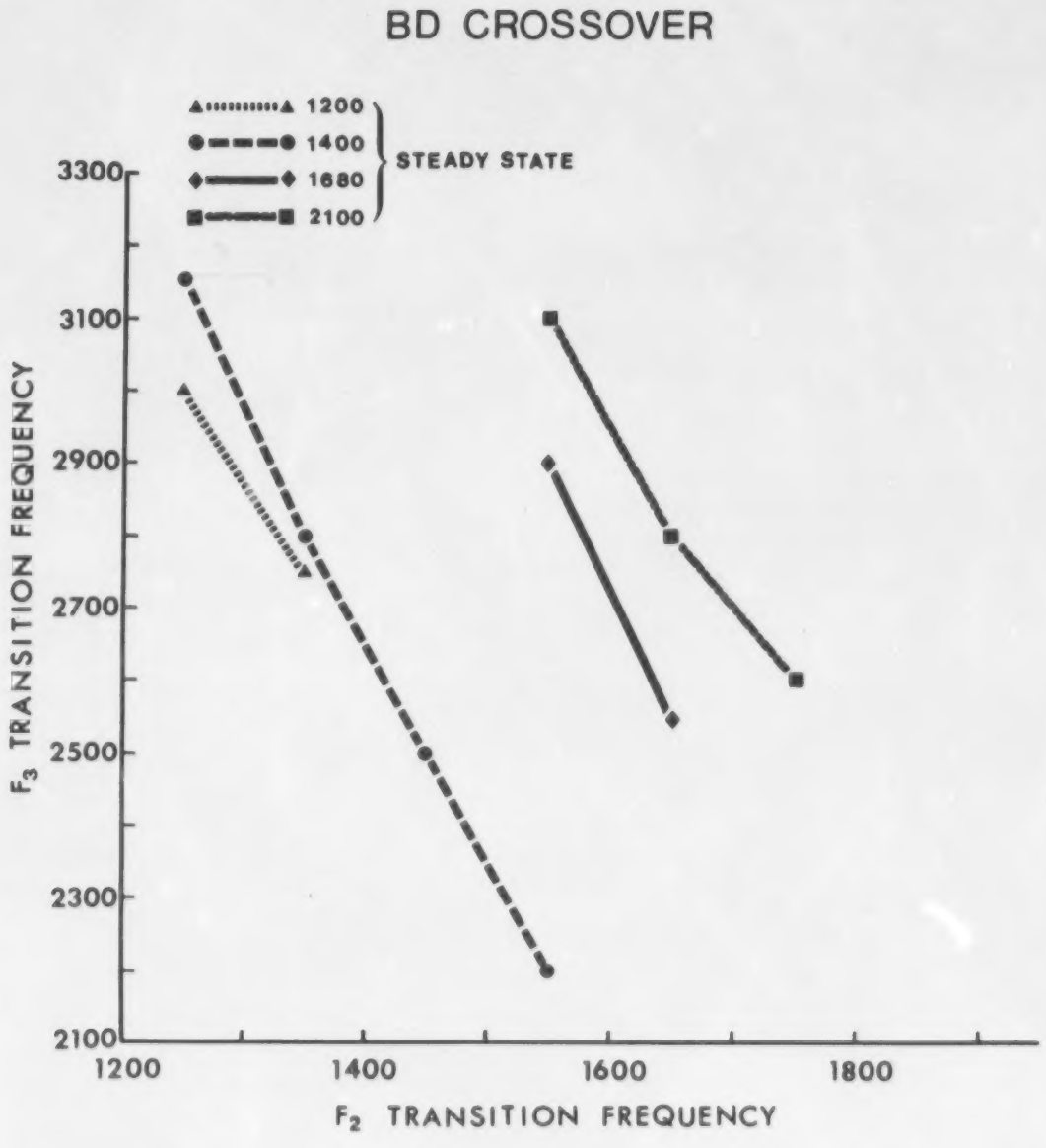

BD CROSSOVER
1200
1400
1680
2100
STEADY STATE
3300
3100
2900
2700
2500
2300
2100
1200
1400
1600
1800
F3 TRANSITION FREQUENCY
F2 TRANSITION FREQUENCY

Figure 2

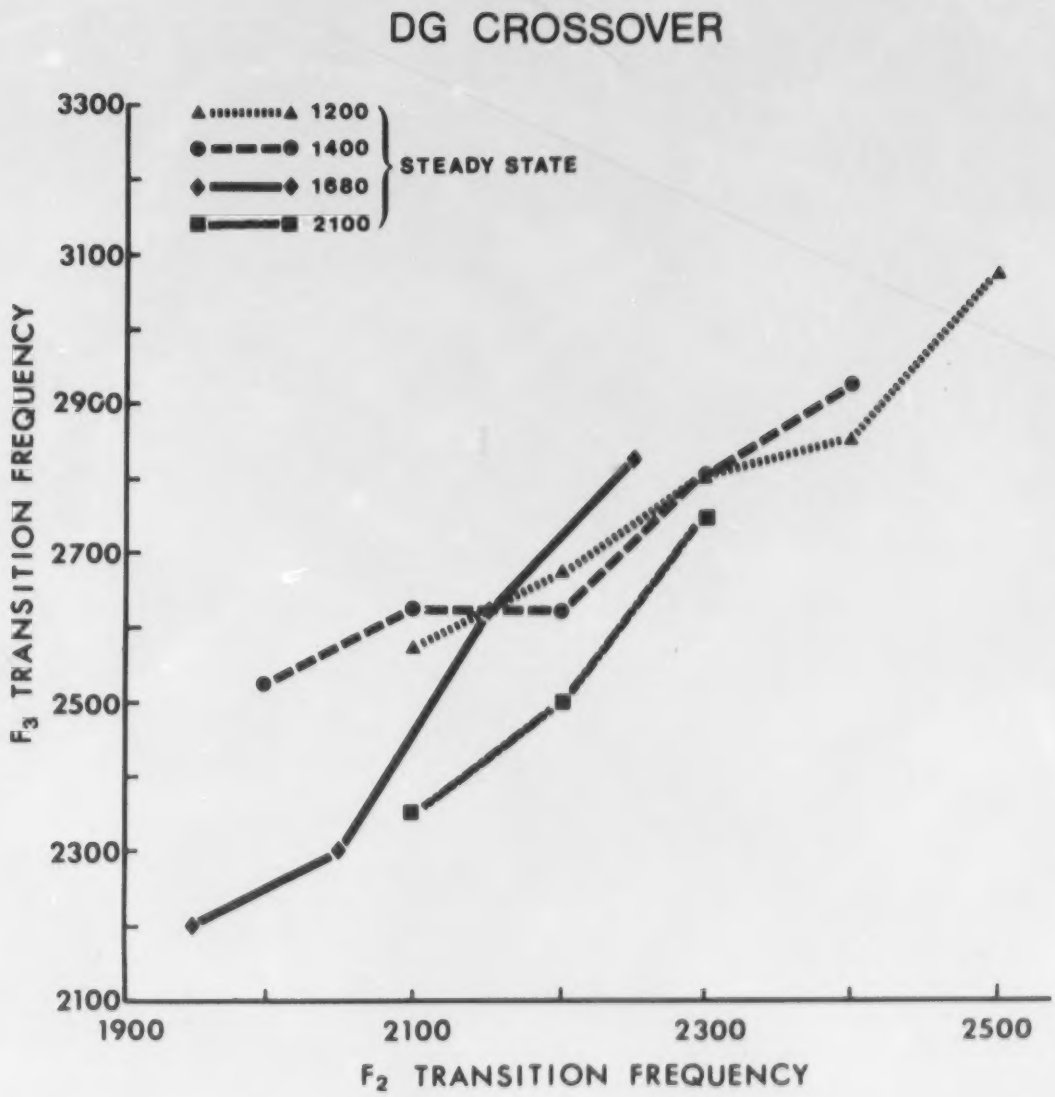

DG CROSSOVER
1200
1400
1680
2100
STEADY STATE
3300
3100
2900
2700
2500
2300
2100
F3 TRANSITION FREQUENCY
1900
2100
2300
2500
F2 TRANSITION FREQUENCY

Figure 3

F2 and F3 engage in strikingly linear trading relations, particularly for the [b-d] distinction, but for the [d-g] distinction as well.

REFERENCES

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America, 1955, 27, 769-773.

Fant, G. The acoustic theory of speech production. The Hague: Mouton & Co., 1960.

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Effect of third-formant transitions on the perception of the voiced stop consonants. Journal of the Acoustical Society of America, 1958, 30, 122-126.

Hoffman, H. S. Study of some cues in the perception of voiced stop consonants. Journal of the Acoustical Society of America, 1958, 30, 1035-1041.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 1954, 68, 1-13.

TOWARD A THEORY OF APRACTIC SYNDROMES*

J. A. Scott Kelso+ and Betty Tuller+

Abstract. Theory development on human motor behavior has, for the most part, occurred independently of data on pathological movement disorders. This paper is an initial attempt to interface findings from studies of apraxia and those of normal motor behavior in order to formulate a common theoretical framework. Such an integration should further the understanding of the nature of skill acquisition and provide insights into the organization of motor systems. Three theoretical models of movement control are discussed with reference to apractic syndromes. The most commonly accepted view—the hierarchy—involves, for example, linear transitivity and unidirectionality of information flow, properties that render it an inadequate explanation of functional plasticity in the central nervous system. The heterarchy, incorporating reciprocity of function and circular transitivity, is a more likely candidate but cannot regulate the degrees of freedom of the system. Our favored candidate is the coalition model, which embodies heterarchical principles, but in addition, offers a solution to the problems of degrees of freedom and context for motor systems. Evidence is reviewed from apraxia of speech and limbs in terms of a coalitional style of control, and an experimental approach, consonant with coalitional organization, is developed. We promote the claim that an understanding of apractic behavior--and perhaps motor systems in general--will benefit when clinicians and experimenters embrace a theory of context and constraints rather than a theory of commands such as those in vogue.

INTRODUCTION

It is an interesting but perhaps distressing feature of science that two different areas of study, each bearing a strong potential relation to the other, can function independently, each in its own oblivion. Except for one or two isolated cases (e.g., Grimm & Nashner, 1978; Roy, 1978), such a situation appears to exist between those who would seek to understand the motor functions of the central nervous system via investigations of clinical disorders and those who seek to understand the underlying behavioral processes

---

*A preliminary version of this paper was given as an invited presentation to the International Neuropsychology Symposium, Oxford, England, 1978. To appear in Brain and Language.

+Also University of Connecticut, Storrs.

Acknowledgment: This work was supported by NINCDS grant NS-13617, BSRG grant RR-05596 and NIH grant AM-25814. We wish to thank K. S. Harris, K. Poeck and D. Shankweiler for comments on an earlier version of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

involved in the acquisition of skill and the control of movement in normal human populations.

The motor behavior area has undergone a good deal of theoretical development in the last decade (e.g., Adams, 1971; Pew, 1974; Schmidt, 1975; Turvey, 1977), but strangely enough, with a total disregard for pathologies of movement. The latter, neurologists suggest, have important ramifications for comprehending motor learning and the organization of motor systems (e.g., Geschwind, 1975). The so-called apractic syndromes seem particularly relevant in this regard, for they arise as a result of cerebral insults that interfere with the generation and elaboration of voluntary movement. An analysis of apraxia therefore affords a unique opportunity for us to derive important insights into the neural control and coordination of learned movements.

Perhaps a major reason for the absence of a viable interface between the research findings of normal and pathological movement behavior is the absence of a theoretical framework that captures a style of organization appropriate to both. Neither closed-loop nor motor program models--currently popular in the motor behavior area (e.g., Kelso & Stelmach, 1976 for review)--address the problem of controlling a complex system containing multiple degrees of freedom. Thus they are unlikely candidates for an adequate analysis of apractic behavior.

In this paper, we will promote a reconceptualization of apraxia based on a perspective on coordinative movement developed primarily by Soviet theorists (e.g., Bernstein, 1967; Gelfand, Gurfinkel, Fomin, & Tsetlin, 1971), and under both theoretical (e.g., Greene, 1972; Turvey, 1977) and empirical (Kelso, Southard, & Goodman, 1979a, 1979b; Nashner, 1976; Nashner & Grimm, 1978) elaboration in this country. In brief, the style of control that we shall propose is coalitional; namely, one in which apraxia is viewed, not simply as a breakdown in function of the nervous system itself (cf. Brown, 1972; Hécaen, 1968), but rather as a decomposition of the syngergistic relationships that hold between the organism and its environment. Current views of apraxia (and indeed motor behavior in general) are based on hierarchical notions of central nervous system organization. This perspective, we shall argue, offers only a limited conceptual framework for analyzing apractic syndromes, being a partial systems approach that fails to account for some important and intriguing phenomena in apraxia. The first part of this paper will be directed towards substantiating this claim and elaborating a coalitional style of control that for us is the minimal organization possessing functional integrity (Turvey, Shaw, & Mace, 1978). This leads us, in the second part, to a reinterpretation of the nature of apraxia and consequent proposals for experimentation that may yield significant dividends in the quest for understanding how actions are coordinated and controlled.

Hierarchical Styles of Motor Organization

Let us first consider some commonly accepted organizational concepts of the motor system that we feel warrant careful scrutiny and reexamination. This is an often-ignored, but important preliminary step since our conceptualization of pathology is likely to correspond closely to our views of central nervous system (CNS) organization. The notion that the motor system is hierarchically organized has a long history that stems from Hughlings

Jackson's (1976) initial observations and insights; it forms the basis for much of our thinking about the nature of skill and the control of movement. There have been numerous recent expressions of the hierarchical viewpoint (e.g., Bruner, 1973; Connolly, 1977), but the basic idea is explicitly defined in Tinbergen's (1950) description of the nervous system as "...Higher centres controlling a number of centres at a next lower level, each of these in turn controlling a number of lower."

Hierarchical assumptions form the basis for investigations in modern neurophysiology and neurology as well as providing an interpretative backdrop. For example, in answer to the question of where the assembly of "action programs" takes place in the CNS, Brooks (1974) suggests, on anatomical grounds, that we look to the limbic system for the drive to move--and thence to frontal and parietal cortex for formation of the needed association. From there the output would be channeled via the cerebellum and basal ganglia through the ventrolateral nucleus of thalamus to the motor cortex, and then finally to the spinal cord and the muscles. Similarly, in clinical neurology circles, Geschwind (1975) seeks an anatomical (and hierarchical) framework for the analysis of movement organization in response to verbal commands. Since Wernicke's area is responsible for spoken language comprehension, when the subject receives an instruction to execute a right handed movement, "...this order is probably transmitted from Wernicke's area through the lower parietal lobe to the left premotor region. The premotor region, in turn, probably controls the precentral motor cortex, which gives rise to the pyramidal tract--a major pathway for motor control--which sends fibers to the spinal cord, where it activates the nerve cells controlling the muscles" (p. 189). These examples clearly illustrate a style of organization that obeys hierarchical principles (or more accurately, linear chaining). Thus the role each element plays in the anatomical chain is fixed and the ordering of dominance relations is immutable.

More recently, Roy (1978) has related the presumed hierarchical structure of the nervous system to the functional disorders observed in apraxia. At the top of the hierarchy, damage to frontal and parietal-occipital areas of cortex results in the disordered planning and sequencing of motor acts characteristic of ideational and ideomotor apractics (Liepmann, 1920). At the next level, lesions in pre-motor cortex lead to an inability to execute the movement sequence properly (premotor apraxia) even though, according to Roy (1978), the patient is still able to plan the motor activity since the frontal areas and the pathways to premotor cortex are preserved. At the lowest level of the hierarchy comes limb-kinetic apraxia--disruption of individual movements within a sequence--which is thought to occur following trauma to the primary areas of the motor cortex.

In all three of the above examples and in the last particularly, we see an effort to relate the functional organization of the nervous system to its neuroanatomical substrate. But the argument from anatomy does not require presupposing a style of organization that is based on hierarchical principles. Although any characterization of movement must be consonant with anatomical fact, the hierarchical description provides an inadequate representation of both anatomical organization and the functional deficits observed in apractic disturbances. We can buttress this claim on several grounds. Consider first the logic behind attempts to relate apractic behavior to site of brain damage.

Its premises are as follows:

(1) Certain lesions in the central nervous system are associated with certain functional deficits.
(2) The effect(s) of the lesion depends on its locus in a neuroanatomically structured hierarchy, i.e., the higher or lower the lesion, the greater of lesser the degree of behavioral deficit.
(3) Therefore, the functional organization of the system is also hierarchical.

The latter conclusion and consequent claims that data on apractic disturbances "...reinforce the need to incorporate features of hierarchical organization into models of motor skill" (Roy, 1978) are based on a curious tautology. Although we are not arguing that structure and function are unrelated, we submit that the conclusion of a functional hierarchy in apractic, and indeed in normal behavior, is predicated upon the a priori assumption that the motor system is hierarchically organized.

But the notion that the movement control system is hierarchically organized is subject to more serious flaws. While hierarchical organization offers economy of control as its principal quality, it pays for this attribute dearly. Thus in the conventional view, information in the CNS undergoes continuous transformation into progressively less abstract levels of description as it flows through the system from the 'hierarch' to the peripheral musculature.[1] In this manner, the system may use a single degree of freedom on the central level to regulate many degrees of freedom at the periphery. Whilst an efficient means of reducing the degrees of freedom in a system, hierarchical principles of CNS function obviously dictate unidirectionality of information flow (Davis, 1976). Assuming, for example, that two structures, x and y, are arranged in a hierarchy at two different levels, the higher system x will always command y; the opposite cannot occur. Furthermore, if x and y are at the same level in the hierarchy, there can be no cross-talk between them (Turvey, Shaw, & Mace, 1978). In other words, in a hierarchy there is no provision for internal communication within central nervous system networks, yet the evidence for unidirectional pathways is shaky at best (e.g., Brown, 1972; Kelso, 1979; Roland, 1979), while the evidence for such 'feedforward' or 'internal feedback' throughout the CNS is virtually unassailable (see Evarts, 1971; Kelso & Stelmach, 1976, for reviews).

A further, and related problem with hierarchical control is that prestige rests with the highest level in the system, i.e., the so-called executive plan or program.[2] Although this is a tempting approach to conceptualizing apractic disorders (which are, after all, 'higher-order' in nature), it has severe limitations if taken seriously.[3] Hierarchies, by definition, dictate that once the superordinate state is lost (say by brain lesions), all subordinate portions governed by it will be left uncontrolled regardless of whether the role of the superordinate state is viewed as excitatory or inhibitory. The absence of excitation or the release from inhibition would render the system unmanageable. In fact, disruption at any level in a hierarchically organized system affects the functioning of all those below it. This shortcoming arises because the relation, 'is governed by' or 'controls,' follows the principle of linear transitivity in hierarchical networks.

But there are data on apractic patients that illustrate the inadequacy of such a view. For example, the abnormally long stretches of jargon speech associated with one form of Wernicke's aphasia have been explained as due to a relatively isolated, and therefore free-running, Broca's area (Geschwind, 1969). But this notion implies a totally random stringing together of elements that is not only an inaccurate description of the grammatical patterns of fluent aphasics, but does not explain why the initial phrase of sentences is often produced correctly (Buckingham & Kertesz, 1974; see also Buckingham, 1979). Similarly, damage to frontal cortex--which typically is assigned superordinate status in the planning of actions (e.g., Luria, 1973; Milner, 1964)--should invariably lead to gross disruption of motor activity, if the hierarchical model is valid. Indeed one could think of few better tests of hierarchical organization than to damage the structure thought to contain the 'hierarch'; motor behavior should disintegrate. But this is not the case: So-called 'habitual skills' (Luria, 1966; Roy, 1978), while rarely produced in response to clinicians' requests, are effectively carried out in the proper context. This finding demands an adequate explanation, but for the time being let us emphasize how damaging it is for hierarchical conceptualizations of motor skill in general, and for apraxia in particular. To preserve the theory, the hierarchist must provide some rationalization for why damage to the structure involved in planning all acts (in this case the frontal or parietal-occipital area) has different effects on some acts than on others. At a minimum, the theorist must provide some basis for distinguishing 'planned' and 'voluntary' acts from those that are 'habitual' or 'automatic' (cf. DeRenzi, Pieczuro, & Vignolo, 1966). More important, we must understand why certain acts may be performed 'given the proper context' (e.g., Roy, 1978; see also Luria, 1966, 1973). Such a rationale does not exist, nor, from our perspective, can it exist within a hierarchical framework. The language of hierarchical systems is one of command, not context: If the motor system is hierarchically organized, then it is, we would argue, context independent.

The broader message of this analysis, then, is that hierarchies provide no means for explaining the functional plasticity in biological systems. Apractic deficits, for example, as pointed out by Geschwind (1975), are often difficult to detect because of this very factor. Alternative pathways not normally used can be brought into operation following insults to primary brain mechanisms. Conventional (hierarchical) accounts of apraxia assume plasticity. But assumed plasticity is incompatible with hierarchical organization because the ability to recruit reserve or back-up structures violates the linear transitivity principle, instead favoring circular transitivity (Turvey et al., 1978). Circular transitivity is a characteristic of heterarchical organization (McCulloch, 1945), a topic to which we briefly turn.

Heterarchical Styles of Organization

Heterarchies embody many features that contrast directly with hierarchies: Rather than ascribing control to a hierarch, heterarchies are characterized by reciprocity of function. In neurology circles, the concept of reciprocity is similar to what Luria (1966) termed "functional pluripotentialism." Thus, in a heterarchical style of organization, a system--by virtue of the reciprocal interconnections amongst its elements--is allowed to assume a variety of roles or functions depending on task demands. Conversely, a particular function may be manifested in a variety of structures; there is no

compartmentalization of function in a heterarchy. Accordingly, control may shift--as a consequence of such distributed function capability--to the source of the most important information. These features of heterarchical organization--reciprocity and distributed function--enable a system to exploit redundancy. Duplication of function and the presence of extensive reciprocal interconnections thus reduce the vulnerability of the system to potential insults, and help preserve its behavioral stability.

If heterarchical organizations closely approximate a realistic characterization of apractic deficits, then attempts to relate functional disorders to locations in the brain or to loss of connections between controlling centers are obviously questionable at best. We can only echo the remarks of Grimm and Nashner (1978) in this regard in their discussion of neurological deficits: "...dimensions of the deficit represent the best mix of systems remaining to participate [emphasis ours]. The characterization of the remanent systems, their redundancy [emphasis theirs] and the limitations they impose on performance are necessary before making a functional correlation between a lesion and a motor disturbance."

The above statement nicely captures the heterarchical style of organization and shifts the emphasis away from the common (and we believe ill-founded) preoccupation with linking function directly to specific structures. But this is not to say that a heterarchical style of organization is an entirely satisfactory conceptual framework for understanding apractic deficits. Although offering a well-motivated rejection of the so-called laws of hierarchical structure (Luria, 1973)--unidirectionality of information flow, centralization of control and compartmentalization of function--all of which, strictly speaking, defy the emergence of functional plasticity, heterarchies bring with them a new set of problems. Perhaps the major one is that the heterarchy--by virtue of its free dominance capability and the fact that locus of control is free to reside anywhere in the system--is too flexible. Thus, while a hierarchy is an effective solution to regulating the degrees of freedom of a system (admittedly with serious consequences), a heterarchy has the problem of managing a potentially infinite number of degrees of freedom.

The degrees of freedom problem is not trivial (cf. Bernstein, 1967), belonging as it does to the class of "non-deterministic polynomial-time complete" problems (Lewis & Papdimetrios, 1978). More simply, the time necessary to regulate a set of independent variables increases as an exponential function of the number of variables to be regulated. Thus, for any living system, the cost (in time) of controlling a large number of degrees of freedom would outweigh the benefits of heterarchical organization. If coordinated movement is to follow heterarchical principles, the number of degrees of freedom to be controlled individually must be reduced. The question arises as to how this may be accomplished. One possibility is that a reduction of the degrees of freedom occurs when a set of variables are linked to form self-regulating autonomous subsystems (cf. Greene, 1972). Wherever the locus of control at any given moment, regulation of the entire subsystem entails only one degree of freedom. Moreover, it makes no sense for variables to be randomly linked to form biologically and behaviorally irrelevant subsystems. Rather there must be a principled basis for constraining variables into appropriate functional units. A good candidate from which such constraints may arise, and one that is motivated by descriptions of apraxia, is the

situational context within which an act is performed. Given this hypothesis, the traditional dichotomy between "habitual" or "automatic" acts and "planned" or "voluntary" acts becomes less tenable. For us then, understanding how a heterarchical system operates entails understanding how a system may be contextually constrained. We feel that the issue of context has been skirted too long in explanations of apractic behavior. What follows is an attempt to conceptualize the notion of 'context of constraint' as it applies to the functioning of the nervous system (or, more appropriately, animals and humans), and to promote an experimental approach that is consonant with it..

What does it mean for a system to be contextually constrained?

> "The meaning of a particular action cannot be explained by a narrow concentration upon the physical movement in isolation. The meaning is given by the context of the action, or complex of actions, of which it can be observed to form a part. Precisely the same physical movements may have quite different meanings, i.e., it may be different actions in different contexts" (Best, 1978).

A popular way of conceptualizing 'context of constraint' is in terms of the activation of a motor image or plan (which itself may be either hierarchical or heterarchical) in the rather restricted sense of a stimulus activating a response. Geschwind (1975), for example, considers a verbal command as an inadequate stimulus for a patient with destruction of the anterior four-fifths of the corpus callosum, in that the experimenter's verbal command cannot reach the patient's right hemisphere. Hence, the verbal stimulus cannot initiate correct responses by the patient's left limbs. In contrast, an object placed in the patient's left hand constitutes a visual stimulus to the right hemisphere, which can evoke the correct movement response. In both cases the movement is considered functionally equivalent regardless of whether it occurs as the response to a verbal or a visual stimulus. By considering verbal commands and situational context as stimuli for functionally equivalent movements, motor apraxias, so it seems, can be understood as a breakdown of the stimulus-response relations that normally hold.

Notice that in this view the relationship between the object (the stimulus) and the actor (the response) is not truly interactive but rather is unidirectional, being characterized by an immutable, hierarchical dominance relation. In order to be effective, the stimulus must activate the response via pathways that are responsible for the interpretation of verbal or visual information or by the motor system itself.

We wish to support an alternative theoretical perspective--based on work by Bransford, McCarrell, Franks, and Nitsch (1977)--in which a movement (or any event) is not defined independently of the context in which it occurs. In this view, in response to questions like "Who wants to go with me?", "How many oranges do you have?" and "How high can you reach?", the gesture of an outstretched hand with all five fingers extended has very different meanings. In short, the hand gesture is not functionally equivalent in different contexts.

In our view, the significance of a movement, and its functional role, are integral to the process of linking free variables into coordinated subsystems. Just as the situational context provides boundary conditions or constraints on the possible meaning of the movement, so also do the possible meanings of the movement provide boundary conditions on the movement's dynamic forms. We will examine this notion in more detail later.

Coalitional styles of organization

Our view of organization, in which actions cannot be considered independently of their context, is captured by what may be termed a coalition (see Turvey, Shaw, & Mace, 1978, for detailed discussion). We have gone to some length in attempting to establish that heterarchical and hierarchical notions consider, in effect, only part of the total system that defines the movement. In contrast, a coalitional framework stresses the mutual compatibility or fit between the individual and the environment. Whilst a coalitional style of control embodies the advantageous characteristics of heterarchies--namely, free dominance, reciprocity, and distributed function--it possesses the additional control advantage of effectively reducing the degrees of freedom of the system.5 Thus, unlike heterarchies where environmental variables are potentially indifferent to the organism and vice versa (hence magnifying the degrees of freedom problem), in coalitions the environment is just as thoroughly organized as the organism and is specific to it (Gibson, 1977; Turvey et al., 1978). Thus, as Turvey et al. (1978) point out, neither member of the synergy is properly constrained without the other, nor may the total system be defined without their closure. From our perspective, then, reduction of the degrees of freedom is accomplished by the contextual framework that operates as a constraint on possible movements. Accordingly, the interaction between the individual and the context or environment must be an adaptive one whose fit is functionally defined by the particular behavioral goal. As a consequence, the significance of this interaction must be an important variable in the coalitional system. If we are correct in claiming that a coalition represents the minimal organization that possesses functional integrity, then apractic deficits may be more properly viewed as a breakdown in the synergistic relationship between the individual and the environment as defined by the behavioral goal.

To summarize, we view the role of informational support or context in a system as providing boundary constraints on the specifics of an action. Our definition of context is very broad and may be applied to both coarse-grain and fine-grain analyses of the nature of control. Accordingly, the significance of a movement, as well as the specifics of the movement, are a function of the coordinative relationship between any particular movement and a set of contextual boundary constraints. Verbal commands, imitation, and even object use, in our view, leave too many degrees of freedom unconstrained. Context, defined globally and locally, provides boundary conditions that specify exactly how the degrees of freedom of meaning and movement must be constrained. Let us illustrate how this may be the case.

Various forms of apraxia (e.g., ideational, ideomotor, constructional) may be characterized broadly as disorders in which the meanings of objects and events are disrupted. We have seen that the meaning of an act in the absence of an appropriate contextual framework is quite different from the meaning of

an act embedded in a particular context, even though the kinematic details may be superficially the same and analyzed as such by the clinician/experimenter. The kinematic sequence exhibited by an apractic patient pretending to hammer a nail in a clinical setting is not functionally equivalent to the (possibly identical) kinematic sequence that occurs when actually hanging a picture. The former is extrinsically specified and applies to only a single part of the system, namely the patient. The latter is specified as a function of the interactions within the total system; the significance, or meaning is an intrinsic feature of the whole act.

EVIDENCE IN SUPPORT OF COALITIONAL CONTROL: TOWARD AN EXPERIMENTAL RE-ANALYSIS OF APRAXIA

Thus far we have argued that given appropriate context the organization of an act is uniquely specified. We would like to examine some experimental evidence supporting the notion that contextual constraints serve to specify precisely the parameters of the motor system. This, we believe, forces a novel, but principled approach to the experimental analysis of apraxia. Key insights into this problem are provided by Belen'kii, Gurfinkel, and Pal'tsev's (1967) demonstration that during the reaction time period for arm movement, the muscles of the trunk and lower limbs undergo a highly patterned and specific series of changes. Note that these muscles are unrelated to those actually involved in the volitional act, but characteristically change before any actual limb movement. Nevertheless, the postural changes that occur depend on the requirements of the intended limb movement, such that those changes specific to raising a leg cannot be identical to those changes specific to raising an arm. The requirements of the intended movement specify the necessary postural adjustments, thereby reducing the number of control decisions required. In other words, the boundary constraints applied to the postural organization minimize what Bernstein (1967) called "the degrees of freedom problem." In the Belen'kii et al. experiment, the complex of postural adjustments is uniquely specified by the nature of the upcoming movement.

We also see from the Belen'kii et al. study that the relationship between postural adjustments and the specific limb movement is not one of immutable, unidirectional dominance. The requirements of the limb movement do not simply impose boundary constraints on the postural mechanisms. While the latter are indeed specific to a movement, e.g., lifting an arm, they in turn preclude the occurrence of a number of other possible activities, e.g., lifting a leg. As Fowler (1977) points out, an individual in this state of "feedforward" is constrained to produce one of a limited class of acts. The postural context provides boundary conditions on what movements are possible while specifics of the intended movement constrain the postural organization. Hence, the relationship between postural adjustments and limb movement must be viewed as at least reciprocal.

It remains to appreciate an additional variable in the relationship between postural adjustments and limb movements before necessarily describing the style of control as coalitional. We have seen how specifics of the intended movement bias the postural system. Nevertheless, the particular movement comprises only part of the contextual framework for postural adjustments: The nature of the support surface must also constrain such modifications. Those changes appropriate for lifting the arm when the individual is

on solid terrain are inappropriate for lifting the arm when in water. Only the coalitional style of control, then, captures the coordinative constraints that exist between the individual, the activity, and the environment.

The so-called 'lower-level' adjustments that normally occur before voluntary movement warrant, in our mind, much more detailed examination. Conceivably, some apractic disturbances result from brain insults that disrupt the complex of supraspinal influences on progressive changes in brainstem and spinal organization. Kots (1977), among others, has examined the changes in spinal organization before and during voluntary movement by testing the excitability of spinal motoneuronal pools. A monosynaptic Hoffman-reflex is elicited by direct electrical stimulation of an afferent nerve. The strength of the reflex provides the information from which one can infer the state of excitability, or gain in the motoneuronal pool. On the basis of empirical evidence, Kots has divided the complex of spinal (and presumably brainstem) changes into three basic processes: pretuning, tuning, and triggering. Pretuning occurs before the signal to move and extends throughout the latent period of the movement. It involves a "background" increase in the reflex excitability of all motoneuron pools and is generally associated with postural adjustments in anticipation of a movement. The pretuning process is associated with supraspinal processes, in that pretuning is absent from the agonist motoneuron pool during the latent period of an elicited reflex movement.

Changes in the spinal apparatus specific to the future movement are described by the processes of tuning and triggering. Approximately 50 to 60 msec before the onset of electromyographic activity in the agonist of the impending movement, there is a smooth and progressive tuning increase in the reflex excitability of the motoneuron pool of the agonist. During the last 25 to 30 msec of this interval, the "fast" motoneurons of the agonist show a sharp increase in reflex excitability simultaneous with a depression of the inhibitory interneuronal system acting on the motoneuronal pool of the future agonist (triggering).

Are impairments of these processes evident in apractic disorders? In patients with cortically localized pyramidal lesions, the background pretuning change in the motoneuron pool of a paralyzed muscle is absent during any attempt to move the paralyzed limb. However, a small background increase in reflex excitability of the same motoneuronal pool occurs during the latent period of voluntary movement of the healthy limb. These observations suggest separate supraspinal mechanisms for changes specific to the limb to be moved and those changes having more global consequences.

An observation by Geschwind (1975) may be relevant here. He observed that axial movements involving bilateral actions of the eyes, neck, or trunk are often executed correctly by apractic patients in response to a verbal command. In contrast, movements of individual limbs, or of the lips, tongue or larynx cannot be produced. He attributes this to the availability of a non-pyramidal motor system which, while capable of elegantly executing axial movements, can only roughly perform discrete movements of individual limbs.

In the current perspective, we do not wish to dichotomize "pyramidal" and "non-pyramidal" motor command systems, but rather prefer to consider the possibility of selective impairment of cerebral influences on spinal and

brainstem organization.

As we see it, the chief function of changes in spinal organization is to provide the postural context in which a limited class of movements may occur. Thus preservation of axial movements by apractic patients, but not movements of individual limbs, may reflect selective impairment of cortical influences on "tuning" and "triggering" changes in spinal organization.

Available evidence does not run counter to the notion of "non-pyramidal" axial movements. The pyramidal pathways are known to have a selective facilitatory effect on "fast" motoneurons (Preston & Whitlock, 1963). "Fast" and "slow" motoneurons show an increase in reflex excitability during tuning, whereas during the triggering process a sharp and selective increase in reflex excitability of only "fast" motoneurons occurs. Hence, the pyramidal pathways may influence the tuning and triggering processes directly, or via non-pyramidal mechanisms. In fact, tuning and triggering are abolished by lesions of the corticospinal system. We believe, however, that classifying movements as the consequences of "pyramidal" or "non-pyramidal" motor commands has less explanatory power than considering movements as arising from pretuning, tuning, or triggering changes in spinal or brainstem organization.

One means of testing this hypothesis is to determine whether the sequence of changes in reflex excitability in the apractic limb during the latent period before movement, as well as the sequence of postural adjustments specific to the movement, differ radically from the changes known to occur in the normal limb. The same technique may be used in apraxia of speech. Reflex changes specific to speech gestures have been recorded from the orbicularis oris muscle during bilabial movements for the syllable /pa/ (Netsell & Abbs, 1975; McClean, 1978; McClean, Folkins, & Larson, 1978). Empirical testing of persons with apraxia of speech may reveal whether or not these reflex changes specific to normal speech gestures are maintained.

If the notion of boundary conditions on a movement is viable, then more extensive biasing, or pretuning (feedforward) adjustments must precede a movement that is not within a context than precede a movement that is more fully specified by its context. Consequently, in patients with apraxia of speech, the first movement of an articulatory sequence may be the most difficult to produce because the feedforward adjustments of brainstem organization that delimit the class of speech movements must be established. In fact, Shankweiler and Harris (1966), Shankweiler, Harris, and Taylor (1968) and Trost and Canter (1974) found that errors in apractic speech occur more frequently on initial sounds than on the same sound in a medial or terminal position. This analysis may also apply to the so-called ideomotor and frontal apraxias, in which patients have difficulty initiating a movement sequence. Moreover, the nature of the biasing adjustment that occurs is a function of the entire act, not simply of the initial segment. Some evidence for this notion is implied by a series of experiments by Sternberg, Monsell, Knoll, and Wright (1978). Normal subjects were provided a list of one or more monosyllabic words, and were told, on the occurrence of a signal light, to begin reciting the list as quickly as possible. Interestingly, the latency from the signal to the onset of the speaker's response increased linearly with the number of words in the list. Although not interpreted in this way by Sternberg et al., these data suggest to us that the necessary feedforward

biasing is a function of the entire sequence of movements. This paradigm may be exploited to explore the progressive biasing of the motor apparatus in patients with apractic disorders. If apractic deficits result from impairment of the feedforward adjustments that would normally precede a movement sequence, then we might not expect patients to maintain a linear relationship between the latency period for initiating a movement sequence and the number of segments in the sequence. If a linear function is obtained, we would expect the slope of the function to be greater than the slope obtained in a normal subject population. In other words, the effect of adding elements to a movement sequence should be more detrimental to apractic patients than to normal subjects. Preliminary evidence for the latter prediction comes from work by Mateer and Kimura (1977) who observed that complex sequences of movements are more likely to uncover apractic deficits than are single motor tasks.

We have suggested that disturbances in 'tuning' may be manifest in certain apractic disorders although the data base is very limited indeed. Moreover, if some forms of apraxia involve a selective disruption of supraspinal biasing influences on lower levels, further predictions are possible as to exactly what specific aspects of motor output should be altered. It has been argued here and elsewhere (Bernstein, 1967; Easton, 1972; cf. Greene, 1972; Kelso, Holt, Kugler, & Turvey, in press; Turvey et al., 1978), that for coordination to occur, the free variables in a complex system must be organized into collectives (Gelfand, Gurfinkel, Tsetlin, & Shik, 1971), or coordinative structures. Such collectives or neuromuscular linkages are created when interneuronal pools in the brainstem and spinal cord are selectively facilitated and inhibited (Bratzlavsky, 1976; Greene, 1972; Gurfinkel et al., 1971). As a consequence of these tunings or biasings, aggregates of neuromuscular variables are constrained to act as functional units, units which may be marshalled temporarily and expressly for the purpose of accomplishing a particular behavioral goal.

Such functional units are thought to govern the spatiotemporal interactions among body parts and may be parameterized in several ways (cf. Boylls, 1975). One form of parameterization is the structural prescription defined as a set of qualitative ratios of activities in the linked muscles, that apply over time, independent of absolute activity levels. On the other hand, the metrical prescription of a coordinative structure specifies the absolute level of activity in linked muscles. The latter may be viewed as a scalar quantity that multiplies the activities of all muscles in the linkage. This view receives strong support from Orlovskii's (1972) data showing that cerebellar stimulation during cat locomotion affects only the magnitude of muscle contraction, leaving unchanged both the period and the timing of periods relative to the cat cycle. This indeed is the principal characteristic of a coordinative structure--namely, when a group of muscles is constrained to act as a unit, some temporal relationship is preserved invariantly over changes in the magnitude of activity (Turvey et al., 1978).

Experimental evidence of invariant temporal relationships in normal movement must act as the reference for judging whether or not temporal relationships in apractic movements are disturbed. Moreover, whether or not the distinction between structural and metrical specification is preserved in normal and apractic movements is open to experimental test (cf. Grimm &

Nashner, 1978). A series of experiments involving the coordinative use of both hands, and hence the concerted working of both hemispheres, illustrates these concepts rather well (Kelso, Southard, & Goodman, 1979a, 1979b). The results suggest that in the coordination of complex movements some temporal relations are preserved over metrical changes. Moreover, the role of context, both fine- and coarse-grain, is apparent.

The question that precipitated the Kelso et al. studies was a simple one. Suppose an individual is asked to produce movements of the upper limbs toward targets, with the movements varying in amplitude and precision requirements, how will she/he respond? Kelso et al. used the well-established relation between movement duration, movement amplitude and target demands (Fitts, 1954) to create such a situation. The key aspect of the Fitts formulation is that movement-time depends on the ratio of movement amplitude to movement precision.

Consider a one-handed movement condition in which the target size is large and the amplitude is small (termed easy) relative to a condition in which the target size is small and the movement amplitude is large (termed difficult). Movement time in the former case will obviously be much shorter in duration. But when these conditions are combined for both hands, the hand producing a short movement to an easy target does not arrive earlier than its more difficult counterpart, as Fitts' Law might predict. Rather, when subjects were asked to strike targets of varying difficulty as quickly and as accurately as possible, they responded with virtually simultaneous movements of the two hands. Moreover, the limb moving to the easy target did not hover over the target or "wait" for its difficult counterpart, but rather moved at quite a different speed. In fact, the limbs under easy-difficult target conditions reached peak velocity and peak acceleration at practically the same time during the movements. Thus, although different spatial demands on the two limbs affected the magnitude of forces produced by each hand, the absolute timing and the segmental durations of parts of the movement--the timing relations of the limb movements--appeared to be an invariant consequence of the two limbs being organized as a functional unit.

An additional experiment explored more directly the influence of contextual constraints in the environment on the dynamics of the functional unit (Kelso, Putnam, & Goodman, Note 1). The subject was required to …ove both hands to separate targets, but one hand was required to move over an obstruction. Under these conditions the movement of the other, unobstructed hand described an arc, showing the influence of the obstruction on the functional unit. Again, the velocity and acceleration patterns of the two limbs were not independent, but rather possessed highly similar characteristics.

We may interpret the contextual constraints provided by the size of the targets, and by the obstruction, as constraining the degrees of freedom of the unit, rather than the individual limb. The synchronous velocity and acceleration patterns of the two limbs suggest a strong interaction between the limbs and is not conducive to an independent programming view. Moreover, the effect of target size is on the functional unit, again suggesting that the coalitional perspective may better represent the style of the control than hierarchies or heterarchies.

It is apparent that timing relationships among limbs normally are preserved over changes in the absolute levels of activity in individual muscles. A close examination of the dynamics of apractic motor output may reveal whether these timing relations are disturbed. Recently, Grimm and Nashner (1978) have made a cogent case for such an approach to what they term "program disorders," of which the apraxias constitute one type. While we are obviously averse to categorizing the apraxias as "program disorders," we are sympathetic with their view that "...all such disorders engender distortions in their structural and metrical prescription which are measurable defects" [emphasis ours, p. 72]. Such remarks remind us also of Geschwind's (1975) recognition that he "...had often accepted as normal, movements that were in fact poorly executed." He goes on to express a lack of surprise "...if some apparently correct response were proved abnormal by more exacting techniques" (p. 194). We feel that the time is ripe for a more precise approach to problems of apraxia within the presently proposed theoretical (coalitional) framework. A detailed movement analysis of apractic disturbances may prove highly informative relative to disruption of spatio-temporal control. We do not feel content with experimental efforts that examine global task situations in the hope of revealing greater performance deficits in apractic than nonapractic subjects (e.g., Heilman, Schwartz, & Geschwind, 1975). While such studies provide interesting facts about the nature of apraxia, product scores such as time on target on a pursuit rotor task tell us nothing about the motor process itself. Distortions in the structural and metrical prescriptions can only be detected by detailed analysis of the dynamics of movement.

Further support for this type of approach comes from recent work on apraxia of speech in which fiberoptic measurements of velar movements were obtained from an apractic speaker (Itoh, Sasanuma, & Ushijima, 1979). When the intended target phoneme /n/ was replaced by /d/, the velum continued to descend for a period of time, that is, the pattern of velar lowering specific to a nasal sound was preserved. This appears indicative of poor temporal coordination between velar lowering for /n/ and tongue tip movement for alveolar closure. Thus, the phonetic change is not an error in selection of the target sound (or meaning element) but results from a breakdown of the tight temporal patterning of movement of two or more articulators. Kent, Carney, and Severeid (1974), in their cinefluorographic study of tongue and velar movements, provide additional evidence that articulatory movements are organized as functional units, so that for a given rate of speaking the relative timing of movements of the tongue, lips, velum, and jaw is systematically patterned. This timing pattern of tongue, lips, and velar movements for normal production of a nasal sound may be disrupted in the speech of an apractic patient (Itoh, Sasanuma, Hirose, Yoshioka, & Ushijima, 1978).

Disturbances of temporal coordination in apraxia of speech have also been suggested by Freeman, Sands, and Harris (1978) and Sands, Freeman, and Harris (1978) who found that their apractic speaker produced a large number of errors on the voicing dimension. As discussed by Lisker and Abramson (1964, 1967) the voicing contrast results from a tight temporal relationship between laryngeal and upper articulator events. Freeman et al. (1978) measured voice onset time for initial stop productions of an apractic speaker and found that the discrete temporal categories for voiced and voiceless stop consonants produced by normal speakers were not preserved in apractic speech.

It appears that apraxia of speech may be characterized, at least in part, by a disruption of the normally invariant timing relations among articulators. To date, however, there have been no investigations of the effect of metrical changes on apractic speech, or even on apractic movements in general. Metrical changes do not exist in isolation but rather specify values for parameters of a structural organization. In the experiments of Kelso et al. (1979a, 1979b), the spatial demands of any one target affected the movement of both limbs. (In the terminology used here, the spatial demands specify values for parameters of the functional unit that includes both limbs.) In another sense, however, the metrical specification provides the background context for the structural organization in that the form of an action and its meaning depend on the metrical specification. Running is not functionally equivalent to walking, although their structural organizations may be identical.

These arguments underscore the broad concept of context that we are proposing. Orlovskii's (1972) data, and indeed the general concept of tuning, show coordinative structures to be the context that gives the supraspinal information meaning. Identical supraspinal signals will have different consequences, or significances, for the individual depending on the background context (coordinative structure organization) of the spinal cord or brainstem (cf. Aizerman & Andreeva, 1968; Greene, 1972). Conversely, the realization of a spinal or brainstem organization is affected by the supraspinal signals. Only a coalitional organization could provide an interpretative backdrop for this view.

As we have seen, the relation between the individual and the environment is specific to the behavioral goal and constrains the movement dynamics. Any analysis in which the motor system is considered to the exclusion of the functional setting must fall short of attaining an adequate specification of apractic coordination. Moreover, the data we have discussed here portray some general principles about how movements are controlled and coordinated and what an account of apraxia must entail.

Obviously our motivation has been to provide a conceptual framework (admittedly incomplete) and to promote ways of approaching the problems of apraxia rather than to point to specific neuronal mechanisms that may be tied (in some, by no means clear, manner) to behavioral deficits. For those who find this displeasing, we would echo Greene's concerns about motor systems research--of which the pathologies constitute a chunk—that the time has come to "know less and to understand more" (in Boylls, 1975, p. 9). An understanding of apractic behavior (and perhaps motor systems in general) will not come, we believe, until the possibility is recognized that any claim about what a given piece of the nervous system tells another piece "...is not the same as what a given piece of the nervous system tells the animal" (Shaw & Turvey, in press). More emphatically, our bias promotes a shift from a theory and language of commands to a theory and language of context.

REFERENCE NOTE

1. Kelso, J. A. S., Putnam, C., & Goodman, D. Visual perturbations reveal a new style of control. Unpublished manuscript.

REFERENCES

Adams, J. A. A closed-loop theory of motor learning. Journal of Motor Behavior, 1971, 3, 111-149.

Aizerman, M. A., & Andreeva, E. A. Simple search mechanism for control of skeletal muscles. Automation and Remote Control, 1968, 29, 452-463.

Belen'kii, V. G., Gurfinkel, V. S., & Pal'tsev, Y. I. Elements of control of voluntary movements. Biophysics, 1967, 12, 154-161.

Bernstein, H. The coordination and regulation of movement. New York: Pergamon, 1967.

Best, D. Meaning in movement. Journal of Human Movement Studies, 1978, 4, 211-222.

Boylls, C. C., Jr. A theory of cerebellar function with applications to locomotion. I. The physiological role of climbing fiber inputs in anterior lobe operation. COINS Technical Report (Computer and Information Science, University of Massachusetts), 1975, 75C-6.

Bransford, J. D., McCarrell, N. S., Franks, J. J., & Nitsch, K. E. Toward unexplaining memory. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing. Hillsdale, N.J.: Erlbaum, 1977.

Bratzlavsky, M. Human brainstem reflexes. In M. Shahani (Ed.), The motor system: Neurophysiology and muscle mechanisms. Amsterdam: Elsevier, 1976.

Brooks, V. B. Some examples of programmed limb movements. Brain Research, 1974, 71, 299-308.

Brown, J. W. Aphasia, apraxia, and agnosia: Clinical and theoretical aspects. Springfield, Ill.: Thomas, 1972.

Brown, J. W. Language representation in the brain. In H. Steklis & M. Raleigh (Eds.), Neurobiology of social communication in primates. New York: Academic Press, 1979.

Bruner, J. Organization of early skilled action. Child Development, 1973, 44, 1-11.

Buckingham, H. W. Explanation in apraxia with consequences for the concept of apraxia of speech. Brain and Language, 1979, 8, 202-226.

Buckingham, H. W., & Kertesz, A. A linguistic analysis of fluent aphasia. Brain and Language, 1974, 1, 43-62.

Connolly, K. The nature of motor skill development. Journal of Human Movement Studies, 1977, 3, 128-143.

Davis, W. J. Organizational concepts in the central motor networks of invertebrates. In R. M. Herman, S. Grillner, P. S. G. Stein, & D. G. Stuart, (Eds.), Neural control of locomotion. New York: Plenum Press, 1976.

DeRenzi, F., Pieczuro, A., & Vignolo, L. A. Oral apraxia and aphasia. Cortex, 1966, 2, 50-73.

Easton, T. A. On the normal use of reflexes. American Scientist, 1972, 60, 591-599.

Evarts, E. V. Feedback and corollary discharge: A merging of the concepts. Neurosciences Research Program Bulletin, 1971, 9, 86-112.

Fitts, P. M. The information complexity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 1954, 47, 381-391.

Fowler, C. A. Timing control in speech production. Bloomington, Ind.: Indiana University Linguistics Club, 1977.

Freeman, F., Sands, E., & Harris, K. S. Temporal coordination of phonation

and articulation in a case of verbal apraxia: A voice onset time study. Brain and Language, 1978, 6, 106-111.

Gelfand, I. M., Gurfinkel, V. S., Fomin, S. V., & Tsetlin, M. L. (Eds.) Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: M.I.T. Press, 1971.

Geschwind, N. Anatomy of the higher functions of the brain. In R. S. Cohen & M. Wartofsky (Eds.), Boston studies in the philosophy of science (Vol. IV). Dordrecht: Reidel, 1969.

Geschwind, N. The apraxias: Neural mechanisms of disorders of learned movements. American Scientist, 1975, 63, 188-195.

Gibson, J. J. The theory of affordances. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing. Hillsdale, N.J.: Erlbaum, 1977.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), Progress in theoretical biology (Vol. 2). New York: Academic Press, 1972.

Grimm, R. J., & Nashner, L. M. Long loop dyscontrol. In J. E. Desmedt (Ed.), Cerebral motor control in man: Long loop mechanisms. Basel: Karger, 1978.

Gurfinkel, V. S., Kots, Y. A., Palt'sev, E. I., & Fel'dman, A. G. The compensation of respiratory disturbances of the erect posture of man as an example of the organization of interarticular interaction. In I. M. Gelfand (Ed.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: M.I.T. Press, 1971.

Hécaen, H. Suggestions of a typology of apraxia. In M. L. Simmel (Ed.), The reach of mind: Essays in memory of Kurt Goldstein. New York: Springer Co., 1968.

Heilman, K., Schwartz, H. D., & Geschwind, N. Defective motor learning in ideomotor apraxia. Neurology, 1975, 2, 1018-1020.

Itoh, M., Sasanuma, S., Hirose, H., Yoshioka, H., & Ushijima, T. Articulatory dynamics in a patient with apraxia of speech: X-ray microbeam observation. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1978, 12, 87-96.

Itoh, M., Sasanuma, S., & Ushijima, T. Velar movements during speech in a patient with apraxia of speech. Brain and Language, 1979, 1, 42-61.

Jackson, J. H. On the affections of speech from disease of the brain. In J. Taylor (Ed.), Selected writings of John Hughlings Jackson. London: Hodder and Stoughton, 1931.

Jackson, J. H. Selected writings of John Hughlin's Jackson (Vol. 1). New York: Basic Books, 1976.

Kelso, J. A. S. Motor-sensory feedback formulations: Are we asking the right questions? The Behavioral and Brain Sciences, 1979, 2, 72-73.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. The concept of coordinative structure as dissipative structure. II. Empirical lines of convergence. To appear in G. E. Stelmach (Ed.), Tutorials in motor behavior. Amsterdam: North Holland, in press.

Kelso, J. A. S., Southard, D. L., & Goodman, D. On the coordination of two-handed movements. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 229-238. (a)

Kelso, J. A. S., Southard, D. L., & Goodman, D. On the nature of human interlimb coordination. Science, 1979, 203, 1029-1031. (b)

Kelso, J. A. S., & Stelmach, G. E. Central and peripheral mechanisms in motor control. In G. E. Stelmach (Ed.), Motor control: Issues and trends. New York: Academic Press, 1976, 1-40.

Kent, R. D., Carney, P. J., & Severeid, L. R. Velar movement and timing: Evaluation of a model for binary control. Journal of Speech and Hearing Research, 1974, 17, 470-488.

Koestler, A. Beyond atomism and holism. The concept of the holon. In A. Koestler & J. R. Smythies (Eds.), Beyond reductionism. Boston: Beacon Press, 1969.

Kots, Ya. M. The organization of voluntary movement. New York: Plenum, 1977.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. Coordination and control in naturally developing systems. In J. A. S. Kelso & J. C. Clark (Eds.), The development of human movement control. London: John Wiley, in press.

Lewis, H. R., & Papdimetrios, C. H. The efficiency of algorithms. Scientific American, 1978, 238, 96-109.

Liepmann, H. Apraxie. Ergon des ges Med., 1920, 1, 516-543.

Lisker, L., & Abramson, A. A cross-language study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.

Lisker, L., & Abramson, A. Some effects of context on voice onset time in English stops. Language and Speech, 1967, 10, 1-28.

Luria, A. P. Higher cortical functions in man. New York: Basic Books, 1966.

Luria, A. P. The working brain. New York: Basic Books, 1973.

Mateer, C., & Kimura, D. Impairment of nonverbal oral movements in aphasia. Brain and Language, 1977, 4, 262-276.

McClean, M. Variation in perioral reflex amplitude prior to lip muscle contraction for speech. Journal of Speech and Hearing Research, 1978, 21, 276-284.

McClean, M. D., Folkins, J. W., & Larsen, C. R. The role of the perioral reflex in lip motor control for speech. Brain and Language, 1979, 7, 42-61.

McCulloch, W. S. A heterarchy of values determined by the topology of nervous nets. Bulletin of Mathematical Physics, 1945, 7, 89-93.

Milner, B. Some effects of frontal lobectomy in man. In J. M. Warren & K. Akert (Eds.), The frontal granular cortex and behavior. New York: McGraw-Hill, 1964.

Nashner, L. M. Adapting reflexes controlling the human posture. Experimental Brain Research, 1976, 26, 59-72.

Nashner, L. M., & Grimm, R. J. Analysis of multiloop dyscontrol in standing cerebellar patients. In J. E. Desmedt (Ed.), Cerebral motor control in man. Basel: Karger, 1978, 300-319.

Netsell, R., & Abbs, J. Modulations of perioral reflex sensitivity during speech movements. Journal of the Acoustical Society of America, 1975, 58, Suppl. S41.

Orlovskii, G. N. The effect of different descending systems on flexor and extensor activity during locomotion. Brain Research, 1972, 40, 359-371.

Pew, R. W. Human perceptual-motor performance. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. New York: Erlbaum, 1974.

Preston, J. B., & Whitlock, D. G. A comparison of motor cortex effect on slow and fast muscle innervation in the monkey. Experimental Neurology, 1963, 7, 327-341.

Roland, P. E. Continuing commentary on sensory feedback to the cerebral cortex during voluntary movement in man. The Behavioral and Brain Sciences, 1979, 2, 307-312.

Roy, E. A. Apraxia: A new look at an old syndrome. Journal of Human Movement Studies, 1978, 4, 191-210.

Sands, E., Freeman, F., & Harris, K. S. Progressive changes in articulatory patterns in verbal apraxia: A longitudinal case study. Brain and Language, 1978, 6, 97-105.

Schmidt, R. A. A schema theory of discrete motor skill learning. Psychological Review, 1975, 82, 225-260.

Shankweiler, D. P., & Harris, K. S. An experimental approach to the problem of articulation in aphasia. Cortex, 1966, 2, 277-292.

Shankweiler, D., Harris, K., & Taylor, M. Electromyographic studies of articulation in aphasia. Archives of Physical Medicine and Rehabilitation, 1968, 49, 1-8.

Shaw, R., & Turvey, M. T. Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & T. Pomerantz (Eds.), Perceptual organization. Hillsdale, N.J.: Erlbaum, in press.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978.

Tinbergen, N. The hierarchical organization of nervous mechanism underlying instinctive behavior. Symposia of the Society of Experimental Biology, 1950, 4, 305-312.

Trost, J. E., & Canter, G. J. Apraxia of speech in patients with Broca's Aphasia: A study of phoneme production accuracy and error patterns. Brain and Language, 1974, 1, 63-79.

Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing. Hillsdale, N.J.: Erlbaum, 1977, 211-266.

Turvey, M. T., Shaw, R. E., & Mace, W. Issues in the theory of action. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Erlbaum, 1978.

FOOTNOTES

[1]In fairness to Hughlings Jackson, we should emphasize that he was very sensitive indeed to the problem of assuming a simple translation between psychical and physical states and vice versa. The more recent neurological literature--with occasional exceptions (cf. Brown, 1979; Buckingham, 1979)--hardly reflects Jackson's cautions (e.g., Jackson, 1878, in Taylor, 1931, p. 156).

[2]We do not feel it necessary to cite references in support of this statement. It is perhaps the most dominant feature of modern motor control theory. Elsewhere we have raised serious questions as to the viability of the concept in natural as opposed to artificial (machine) systems (Kelso, Holt, Kugler, & Turvey, in press; Kugler, Kelso, & Turvey, in press).

[3]Of course, it has occurred to us that many who use the term 'hierarchy' to reflect the functional organization of the CNS do not take the concept seriously. If this is so, then our refutation of the notion--as a scientific enterprise--is probably not to be considered seriously either.

[4]This language may antagonize some readers. But the language of hierarchical systems, it seems to us, presupposes one-to-one mappings between structure and function. Thus when it is written "...The premotor region, in turn, probably controls the precentral motor cortex" (Geschwind, 1975, p. 189), we can only assume that the premotor cortex contains a (or even the) controller.

[5]There is a fourth feature of heterarchies embodied by the concept of coalition that is beyond the scope of this discussion, namely, the notion of emergent properties. This is a general but oft-ignored characteristic of complex systems that illustrates the transparency of views that implicitly or explicitly postulate one-to-one mappings between structure and function. As Davis (1976) points out in his discussion of neural network properties, a novel function may reflect the interaction of cells with different properties. For example, oscillation in a motor network may emerge as a property of the system even though no single neuron within the network has the capacity of endogenous oscillation. The broader point, of course, is that no analysis of the system, in terms of its parts or the arrangements between them, will account for emergent phenomena occurring at high levels of organization (cf. Koestler, 1969). Such a view, which we endorse, forces consideration of the total system (expressed as a coalition) in any analysis of movement disorders.

COMPARISON OF PARAMETRIC REPRESENTATIONS FOR MONOSYLLABIC WORD RECOGNITION IN CONTINUOUSLY SPOKEN SENTENCES*

Steven B. Davis+ and Paul Mermelstein++

Abstract. Several parametric representations of the acoustic signal were compared as to word recognition performance in a syllable-oriented continuous speech recognition system. The vocabulary included many phonetically similar monosyllabic words, therefore the emphasis was on ability to retain phonetically significant acoustic information in the face of syntactic and duration variations. For each parameter set (based on a mel-frequency cepstrum, a linear frequency cepstrum, a linear prediction cepstrum, a linear prediction spectrum, or a set of reflection coefficients), word templates were generated using an efficient dynamic warping method, and test data were time registered with the templates. A set of ten mel-frequency cepstrum coefficients computed every 6.4 ms resulted in the best performance, namely 96.5% and 95.0% recognition with each of two speakers. The superior performance of the mel-frequency cepstrum coefficients may be attributed to the fact that they better represent the perceptually relevant aspects of the short-term speech spectrum.

1. INTRODUCTION

The selection of the best parametric representation of acoustic data is an important task in the design of any speech recognition system. The usual objectives in selecting a representation are to compress the speech data by eliminating information not pertinent to the phonetic analysis of the data and to enhance those aspects of the signal that contribute significantly to the detection of phonetic differences. When a significant amount of reference information is stored, such as different speakers' productions of the vocabulary, compact storage of the information becomes an important practical consideration.

---

*To appear in IEEE Transactions on Acoustics, Speech and Signal Processing.

+Now at Signal Technology, Inc., 15 W. De La Guerra St., Santa Barbara, CA 93101

++Now at Bell-Northern Research and INRS-Telecommunications, University of Quebec, 3, Place du Commerce, Nuns' Island, Verdun, Quebec, Canada H3E 1H6

Acknowledgement. This material is based upon work supported by NSF Grant BNS 7682023 to Haskins Laboratories. Mrs. Frank Cooper and Patrick Nye participated in numerous discussions of the experimental program, and their contribution is greatly appreciated.

[HASKINS LABORATORIES: Status Report on Speech Research SR-61 (1980)]

The choice of a basic phonetic segment bears closely on the representation problem because the decision to identify an unknown segment with a reference category is based on the parameters within the entire segment. The number of different reference segments is generally smaller than the number of possible unknown segments, and therefore the step of identifying an unknown with a reference entails a significant loss of information. One can minimize the loss of useful information by examining different parametric representations in the framework of the specific recognition system under consideration. However, since the choice of a segment is so basic to the decision as to what acoustic information is useful, the result of such a comparative examination of different representations is directly applicable only to the specific recognition system, and generalization to differently organized systems may not be warranted.

Fujimura (1975) and Mermelstein (1975b) discussed in detail the rationale for use of syllable-sized segments in the recognition of continuous speech. The goal of the experiments reported here was to select an acoustic representation most appropriate for the recognition of such segments. The methods used to evaluate the representations were open testing, where the training data and test data were independently derived, and closed testing, where these data sets were identical. In each case, the same speaker produced both the reference and test data, which included the same words in a variety of different syntactic contexts. Although variation between speakers is an important problem in its own right, attention is focused here on speaker dependent representations to restrict the different sources of variation in the acoustic data.

White and Neely (1976) showed that the choice of parametric representations significantly affects the recognition results in an isolated word recognition system. Two of the best representations they explored were a 20-channel bandpass filtering approach using a Chebychev norm on the logarithm of the filter energies as a similarity measure, and a linear prediction coding approach using a linear prediction residual (Itakura, 1975) as a similarity measure. From the similarity of the corresponding results, they concluded that bandpass filtering and linear prediction were essentially equivalent when used with a dynamic programming time alignment method. However, that result may be due to the absence of phonetically similar words in the test vocabulary.

Because of the known variation of the ear's critical bandwidths with frequency (Feldtkeller & Zwicker, 1956; Schroeder, 1977), filters spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the phonetically important characteristics of speech. Pols (1977) showed that the first six eigenvectors of the covariance matrix for Dutch vowels of three speakers, expressed in terms of 17 such filter energies, accounted for 91.8% of the total variance. The direction cosines of his eigenvectors were very similar to a cosine series expansion on the filter energies. Additional eigenvectors showed an increasing number of oscillations of their direction cosines with respect to their original energies. This result suggested that a compact representation would be provided by a set of mel-frequency cepstrum coefficients. These cepstrum coefficients are the result of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale.[1]

A preliminary experiment (Mermelstein, 1976) showed that the cepstrum coefficients were useful for representing consonantal information as well. Four speakers produced 12 phonetically similar words, namely "stick," "sick," "skit," "spit," "sit," "slit," "strip," "scrip," "skip," "skid," "spick," and "slid." A representation using only two cepstrum coefficients resulted in 96% correct recognition of this vocabulary. Given these encouraging results, it became important to verify the power of the mel-frequency cepstrum representation by comparing it to a number of other commonly used representations in a recognition framework where the other variables, including vocabulary, are kept constant.

This paper compares the performance of different acoustic representations in a continuous speech recognition system based on syllabic units. The next section describes the organization of the recognition system, the selection of the speech data, and the different parametric representations. The following section describes the method for generating the acoustic templates for each word by use of a dynamic warping time alignment procedure. Finally, the results obtained with the various representations are listed and discussed from the point of view of completeness in representing the necessary acoustic information.

2. The Experimental Framework

A rather simple speech recognition framework served as the testbed to evaluate the various acoustic representations. Lexical information was utilized in the form of a list of possible words and their corresponding acoustic templates, and these words were assumed to occur with equal likelihood. No syntactic or semantic information was utilized. If such information had been present, it could have been used to restrict the number of admissible lexical hypotheses or assign unequal probabilites to them. Thus, in practice, instead of matching hypotheses to the entire vocabulary, the number of lexical hypotheses that one evaluates may be reduced to a much smaller number. This reduction would cause many of the hypotheses phonetically similar to the target word to be eliminated from consideration. Thus the high phonetic confusability of the test data may have resulted in a test environment that is more rigorous than would be encountered in practice.

2.1 Selection of Corpus

The performance of continuous speech recognition systems is determined by a number of distinct sources of acoustic variability, including speaker characteristics, speaking rate, syntax, communication environment and recording and/or transmission conditions. The focus of the current experiments is acoustic recognition in the face of variability induced in words of the same speaker by variation of the surrounding words and by syntactic position. The use of a separate reference template for each different syntactic environment which a word might occupy would require exorbitant amounts of storage and training data. Thus an important practical requirement is to generate reference templates without regard to the syntactic position of the word. To avoid the problem of automatically segmenting complex consonantal clusters, the corpus was composed of monosyllabic target words that were semantically acceptable in a number of different positions in a given syntactic context. Since acoustic variation due to different speakers is a distinctly separate

problem (Rabiner, 1978), it was considered advisable to restrict the scope of these initial experiments by using only speaker dependent templates. That is, both reference and test data were produced by the same speaker.

The sentences were read clearly in a quiet environment and recorded using a high quality microphone. These recording conditions were selected to establish the best performance level that one could expect the recognition system to attain. Environments with higher ambient noise, which may be encountered in a practical speech input situation, would undoubtedly detract from the clarity of the acoustic information and therefore result in lower performance.

The speech data comprised 52 different CVC words from two male speakers (DZ and LL), and a total of 169 tokens were collected from 57 distinct sentences (Appendix A). The sentences were read twice by each speaker in recording sessions separated in time by two months (denoted as DZ1, DZ2, LL1 and LL2). Thus the data consisted of a total of 676 syllables. To achieve the required variability, the selected words could be used as both nouns and verbs. For example, "Keep the hope at the bar" and "Bar the keep for the yell" are two sentences that allow syntactic variation but preserve the same overall intonation pattern. All the words examined carried some stress; the unstressed function words were not analyzed. The target words, all CVC's, included 12 distinct vowels, /i, I, e, ɛ, æ, ɔ, ʌ, U, u, ɝ, a, o/, some of which are normally diphthongized in English. Each vowel was represented in at least four different words, and these words manifested differences in both the prevocalic and postvocalic consonants. The consonants comprised simple consonants as well as affricates but no consonantal clusters.

2.2 Segmentation

An automatic segmentation process (Mermelstein, 1975a) was initially considered as one way of delimiting syllable-sized units in continuously spoken text, but any such algorithm performs the segmentation task with a finite probability of error. In particular, weak unstressed function words sometimes appear appended to the adjacent words carrying stronger stress. Additionally, in this study, a boundary point located for an intervocalic consonant with high sonority may not consistently join that consonant to the word of interest. In order to avoid possible interaction between segmentation errors and poor parametric representations, manual segmentation and auditory evaluation was used to accurately delimit the signal corresponding to the target words. The segmentation, as well as the subsequent analysis and recognition, was performed on a PDP-11/45 minicomputer with the Interactive Laboratory System (Pfeifer, 1977).

In systems employing automatic segmentation, the actual recognition rates can be expected to be lower due to the generation of templates from imperfectly delimited words (Mermelstein, 1978). However, there is no reason to believe that segmentation errors would not detract equally from the recognition rates obtained for the various parametric representations.

2.3 Parametric Representations

The parametric representations evaluated in this study may be divided into two groups, those based on the Fourier spectrum and those based on the linear prediction spectrum. The first group comprises the mel-frequency cepstrum coefficients (MFCC) and the linear-frequency cepstrum coefficients (LFCC). The second group includes the linear prediction coefficients (LPC), the reflection coefficients (RC), and the cepstrum coefficients derived from the linear prediction coefficients (LPCC). A Euclidean distance metric was used for all cepstrum parameters, since cepstrum coefficients are derived from an orthogonal basis. This metric was also used for the RC, in view of the lack of an inherent associated distance metric. The LPC were evaluated using the minimum prediction residual distance metric (Itakura, 1975).

Each acoustic signal was lowpass filtered at 5 kHz and sampled at 10 kHz. Fourier spectra or linear prediction spectra were computed for sequential frames 64 points (6.4 ms) or 128 points (12.8 ms) apart. In each case, a 256 point Hamming window was used to select the data points to be analyzed. (A window size of 128 points produced degraded results).

For the MFCC computations, 20 triangular bandpass filters were simulated as shown in Figure 1. The MFCC were computed as

$$MFCC_i = \sum_{k=1}^{20} X_k \cos[i(k - \tfrac{1}{2})\tfrac{\pi}{20}] \quad , \quad i = 1,2,\ldots,M \quad , \tag{1}$$

where M is the number of cepstrum coefficients, and X_k, $k = 1,2,\ldots,20$, represents the log-energy output of the kth filter.

The LFCC were computed from the log-magnitude Discrete Fourier Transform (DFT) directly as

$$LFCC_i = \sum_{k=0}^{K-1} Y_k \cos(\tfrac{\pi i k}{K}) \quad , \quad i = 1,2,\ldots,M \quad , \tag{2}$$

where K is the number of DFT magnitude coefficients Y_k.

The LPC were obtained from a 10th order all-pole approximation to the spectrum of the windowed waveform. The autocorrelation method for evaluation of the linear prediction coefficients was used (Markel & Gray, 1976). The RC were obtained by a transformation of the LPC which is equivalent to matching the inverse of the LPC spectrum with a transfer function spectrum that corresponds to an acoustic tube consisting of ten sections of variable cross-sectional area (Wakita, 1973). The reflection coefficients determine the fraction of energy in a travelling wave that is reflected at each section boundary.

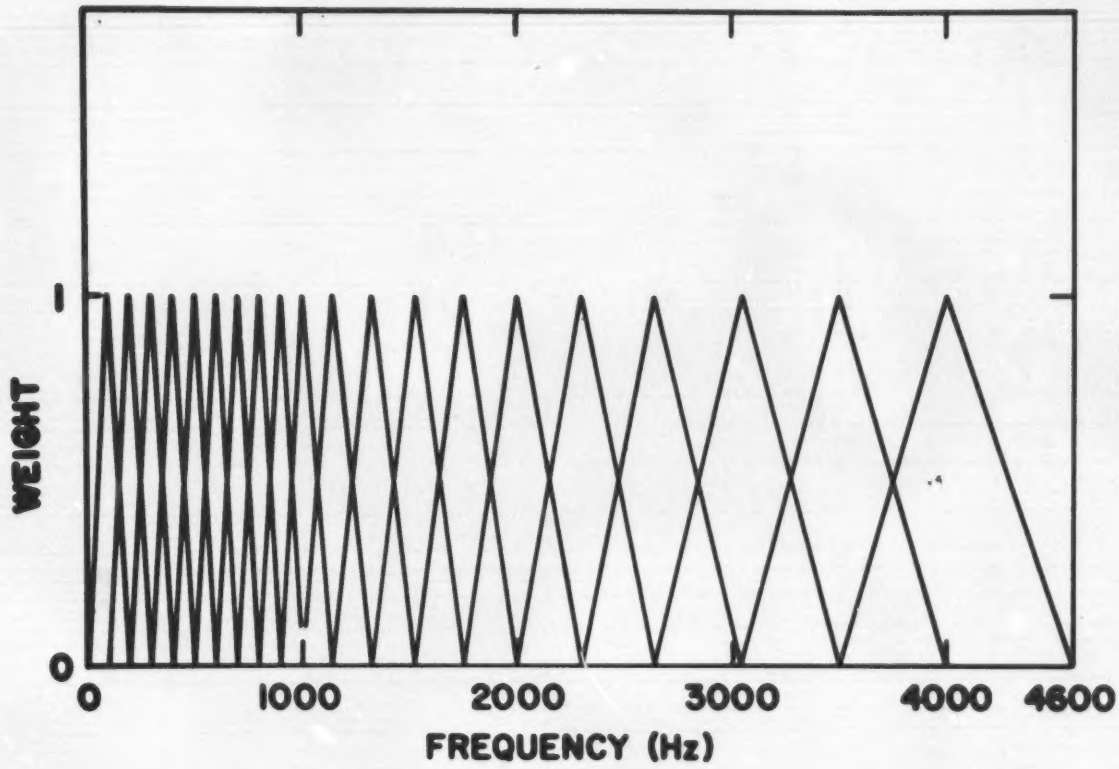


Figure 1: Filters for generating mel-frequency cepstrum coefficients.

The LPCC were obtained from the LPC directly as

$$LPCC_i = LPC_i + \sum_{k=1}^{i-1} \frac{k-i}{i} LPCC_{i-k} LPC_k \quad , \; i=1,2,\ldots,10 \quad . \tag{3}$$

The Itakura metric represents the distance between two spectral frames with optimal (reference) LPC and test $\widehat{LPC}$ as

$$D\,[LPC,\widehat{LPC}] = \log \left| \frac{LPC\ \hat{R}\ LPC^T}{\widehat{LPC}\ \hat{R}\ \widehat{LPC}^T} \right| \quad , \tag{4}$$

where R is the autocorrelation matrix (obtained from the test sample) corresponding to the $\widehat{LPC}$. The metric measures the residual error when the test sample is filtered by the optimal LPC. Because of its asymmetry, the Itakura metric requires specific identification of the reference coefficients (LPC) and the test coefficients ($\widehat{LPC}$). For computational efficiency, the denominator of (4) will be unity if $\hat{R}$ is expressed in unnormalized form. Then if $\hat{r}(n)$ denotes the unnormalized diagonal elements of $\hat{R}$, $r_{LP}(n)$ denotes the unnormalized autocorrelation coefficients from the LPC polynomial, and the logarithm is eliminated, the distance may be expressed as (Gray & Markel, 1976)

$$D[\hat{r},r_{LP}] = \hat{r}(0)r_{LP}(0) + 2 \sum_{i=1}^{10} \hat{r}(i)r_{LP}(i) \quad . \tag{5}$$

3. Generation of Acoustic Templates

The use of templates to represent the acoustic information in reference words allows a significant computation reduction compared to use of the reference tokens themselves. The design of a template generation process is governed by the goal of finding the point in acoustic space that simultaneously minimizes the "distance" to all given reference items. Where the appropriate distance is a linear function of the acoustic variables, this goal can be realized by the use of classic pattern recognition techniques. However, phonetic features are not uniformly distributed across the acoustic data, and therefore perceptually motivated distance measures are nonlinear functions of those data. To avoid the computationally exorbitant procedure of simultaneously minimizing the set of nonlinear distances, templates are incrementally generated by introducing additional acoustic information from each reference word to the partial template formed from the previously used reference words. Given a distance between two tokens, or between a token and a template, the new template can be located along the line whose extent measures that distance. Since only acoustically similar tokens are to be combined into individual templates, one may expect that this procedure will exploit whatever local linearization the space permits.

3.1 Template Generation Algorithms

In one algorithm (Rabiner, 1978), an initial template is chosen as the token whose duration is the closest to the average duration of all tokens representing the same word (Figure 2). Then all remaining tokens are warped to the initial template. The warping is achieved by first using dynamic programming to provide a mapping (or time registration) between any test token and the reference template. Following the notation in Rabiner, Rosenberg, and Levenson (1978), let $T_i(m)$, $0 \leq m \leq M_i$, be a test contour for word replication i with duration M_i, $i=1,2,\ldots,I$, and let $R_1(m) = T_j(m)$ be the initial reference contour, where the duration of the j^{th} token is closest to the average duration. For example, these contours may be vectors of cepstrum coefficients obtained at 10 ms intervals during the word. Then dynamic programming may be used to find mappings $m_i = w_i(n)$, $i=1,2,\ldots,I$, subject to boundary conditions at the endpoints, such that the total distance $D_T(i)$ between test token i and the reference contour is minimal. A distance function D is defined for each pair of points (m,n). Then

$$D_T(i) = \min_{\{w_i(n)\}} \sum_{n=1}^{N} D[R_1(n), T_i(w_i(n))] \quad . \tag{6}$$

With the aid of these mappings, a new reference contour may be defined as

$$R_2(n) = \frac{1}{I} \sum_{i=1}^{I} T[w_i(n)] \quad , \tag{7}$$

and the process is repeated until the distance between the current and previous templates is below some threshold. This procedure is not dependent on the order in which tokens are considered. However, it is computationally expensive to iterate to the final reference contour. Furthermore, there may be cases where there is no convergence (Rabiner, 1978).

A different algorithm can be used for phonetically similar words; this algorithm requires less computation effort and has no convergence problems. Furthermore, the algorithm allows a reference template to be easily updated with an accepted token during verification to allow for word variation over time. In this procedure (Davis, 1979), each successive token is warped with the current template to produce a new template for the next token (Figure 3). For example,

$$\begin{aligned}
R_1(n) &= T_1(n) \quad , \\
R_2(n) &= \frac{1}{2}\,[\,R_1(n) + T_2(w_2(n))] \quad , \\
R_3(n) &= \frac{1}{3}\,[2R_2(n) + T_3(w_3(n))] \quad , \\
&\;\;\vdots \qquad\qquad \vdots \\
R_I(n) &= \frac{1}{I}\,[(I-1)R(I-1,n) + T_I(w_I(n))] \quad .
\end{aligned} \tag{8}$$

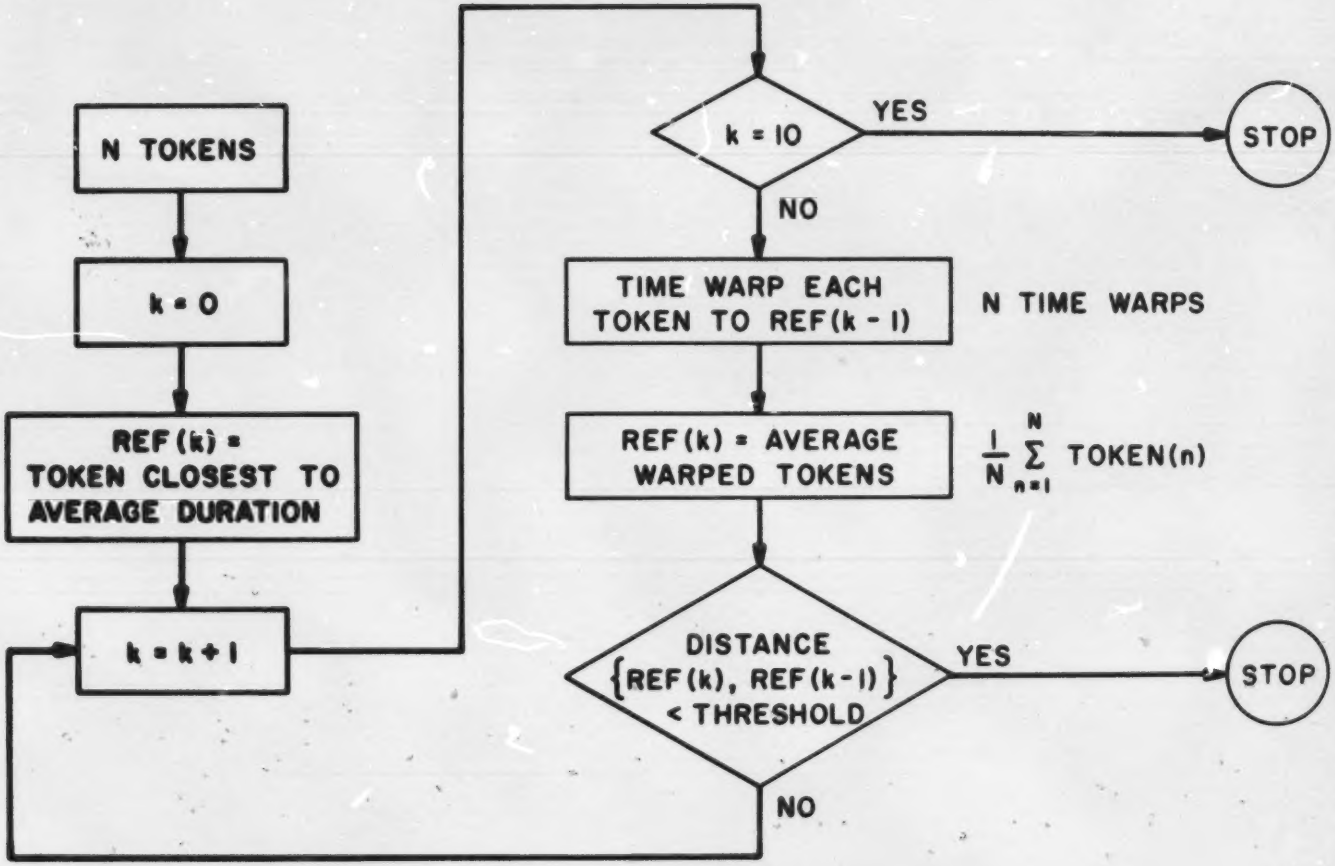


Figure 2: Iterative algorithm for template generation.

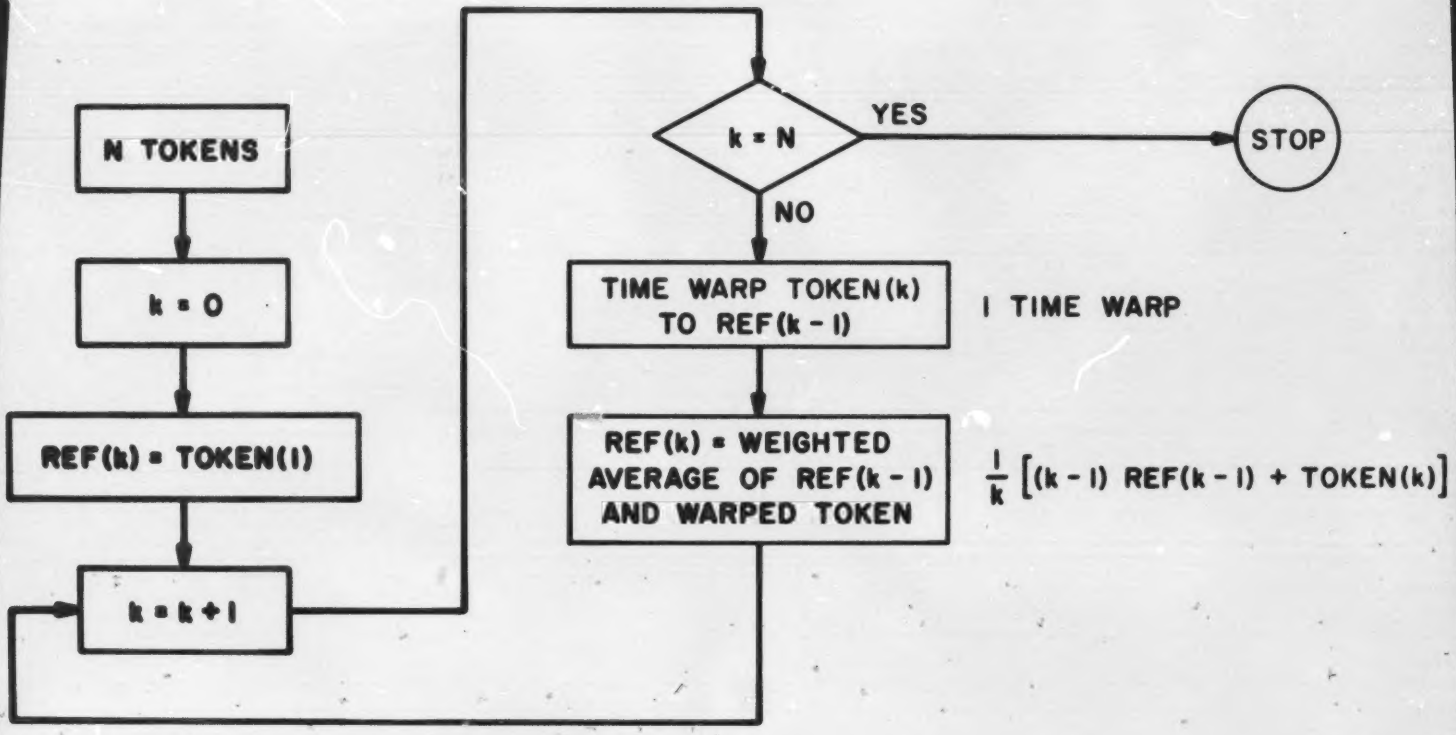


Figure 3: Noniterative algorithm for template generation.

Thus, the process ends with the Ith template.

While this algorithm has computational advantages over the first algorithm, the results become order dependent since the warping is sequential and nonlinear. If the tokens are used in a different order, a different template will result. For tokens obtained from the same speaker and spoken within the same context, order dependence is not a problem. However, for tokens obtained from different syntactic positions, order dependence is potentially a problem. Finally, if different speakers are involved, tokens will be less similar, and the order in which they are taken may greatly affect the final template. If clustering algorithms are used to generate multiple templates for each word (Rabiner, 1978), then each cluster may be viewed as a group in which order dependence may be a consideration.

3.2 Time Alignment

All but one of the parametric distance measures explored are derived from Euclidean functions of parameters pertaining to pairs of time frames. The appropriate time frames are chosen to best align the significant acoustic events in time. Because the segments aligned are monosyllabic words, one can take advantage of a number of well defined acoustic features to guide the alignment procedure. For example, the release of a prevocalic voiced stop or the onset of frication of a postvocalic fricative manifest themselves by means of such acoustic features. The particular alignment procedure used meets these requirements without requiring explicit decisions concerning the nature of the acoustic events.

The alignment operation employed a modified form of the dynamic programming algorithm first applied to spoken words by Velichko and Zagoruyko (1970) and subsequently modified by Bridle and Brown (1974) and Itakura (1975). In view of the intent to use the same algorithm for template generation as for recognition of unknown tokens, a symmetric dynamic programming algorithm was utilized. Sakoe and Chiba (1978) have recently shown that a symmetric dynamic programming algorithm yields better word recognition results than previously used asymmetric forms.

Execution of the algorithm proceeded in two stages (Figure 4). First, the pair of tokens to be compared was time aligned by appending silence to the marked endpoints and linearly shifting the shorter of the pair with respect to the longer to achieve a preliminary distance minimum. Since monosyllabic words generally possess a prominent syllabic peak in energy, this operation ensured that the syllabic peaks were lined up before the nonlinear minimization process was started. Informal evaluation has shown that use of the preliminary alignment procedure yields better results than omitting the procedure or using a linear time warping procedure to equalize the time durations of the tokens. The two tokens, extended by silence where necessary, were then subjected to the dynamic programming search to find an improved distance minimum. The preliminary distance minimum, found as a result of the initial linear time alignment procedure, corresponded to the distance computed along the diagonal of the search space and represented in most cases a good starting point for the subsequent detailed search. Use of this preliminary time alignment, and the additional invocation of a penalty function when the point selected along the dynamic programming path implied unequal time

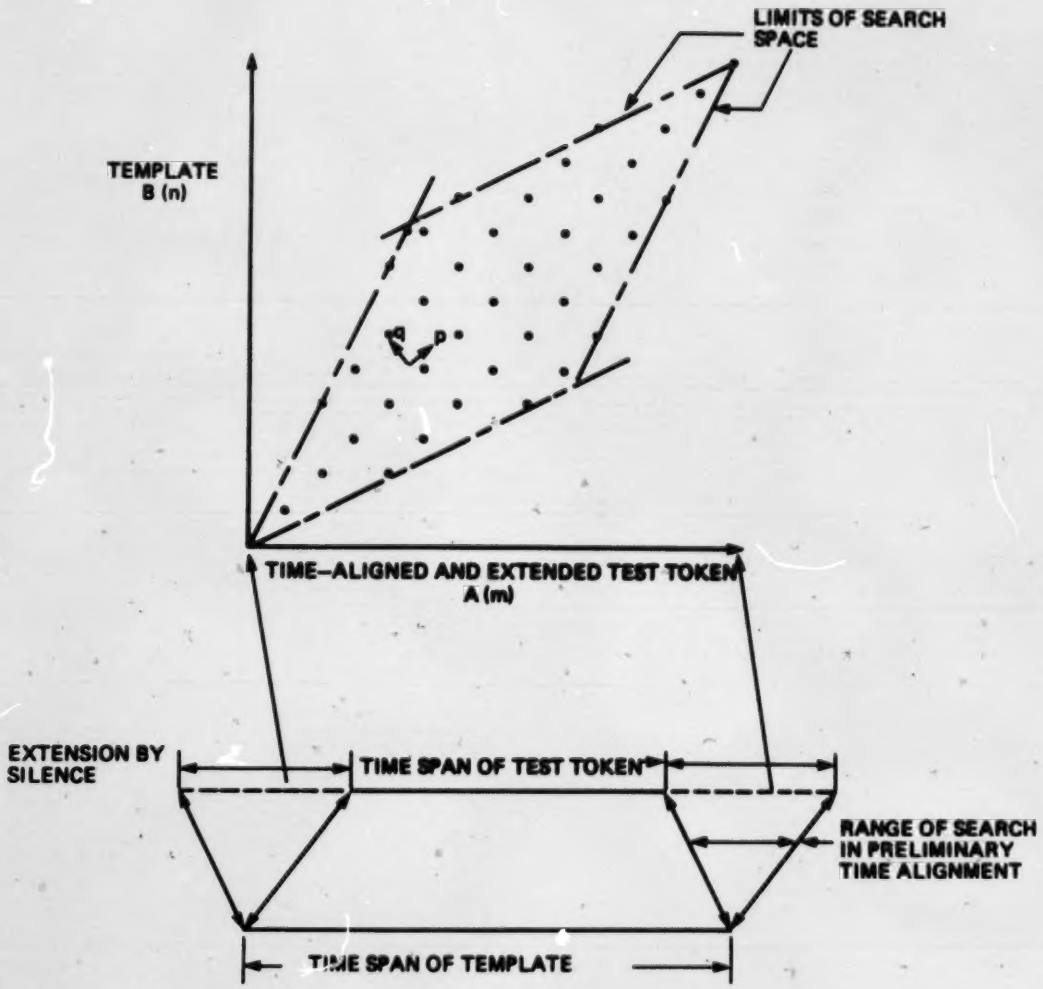

LIMITS OF SEARCH SPACE
TEMPLATE B (n)
TIME–ALIGNED AND EXTENDED TEST TOKEN A (m)
EXTENSION BY SILENCE
TIME SPAN OF TEST TOKEN
RANGE OF SEARCH IN PRELIMINARY TIME ALIGNMENT
TIME SPAN OF TEMPLATE

Figure 4: Dynamic time alignment of speech samples.

increments along the measured data, generally forced the optimum warping path to be near the diagonal, unless prominent acoustic information was present to indicate the contrary. For efficiency in programming, zeros (representing silence) were never really appended to the data, rather, the time shift was retained and used to trigger a modified Euclidean or Itakura distance measure when appropriate.

The use of silence to extend the syllable tokens in the preliminary time alignment, instead of linear time expansion or contraction as implied by asymmetric formulations of the dynamic programming algorithm, requires some justification. The comparison here is among syllable-sized units which generally possess an energy peak near the center regions and lesser energy near the ends. Based on a perceptual model, extension of the tokens by silence is clearly appropriate. Linear time scale changes would obscure equally the more significant duration information in the consonantal regions and less significant duration information in the vocalic regions. Discrimination between words like "pool" and "fool" depends critically on the duration of the prevocalic burst or fricative. The alignment ensures that the prominent vowel regions are lined up before time scale changes in the consonantal regions are examined.

3.3 Dynamic Warping Algorithm

The dynamic warping algorithm serves to estimate the similarity between an unknown token and a reference template. Additionally, it serves to align a reference token with a partial template to ensure that phonetically similar spectral frames are averaged in generating a composite template. Through the preliminary alignment procedure discussed above, the token or template, whichever is shorter, is extended by silence frames on both sides. The resulting multidimensional acoustic representations of the pair of patterns compared can be denoted by $A(m)$, $m = 1,2,\ldots,M$ and $B(n)$, $n = 1,2,\ldots,M$. For each pair of frames $\{A(m), B(n)\}$, a local distance function $D[A,B]$ can be defined for estimating the similarity at point $x'(m,n)$. A change of variables identifies $x'(m,n)$ as $x(p,q)$, where p and q are measured along and normal to the diagonal illustrated in Figure 4. For each position along the diagonal $\{x(p,0),\ 1 \leq p \leq M\}$, points along the normal $\{x(p,q),\ |q| \leq Q(p)\}$ are analyzed, where the search space is limited by $|q| \leq Q(p)$. The $Q(p)$ define a region in the grid area delimited by lines with slopes 1/2 and 2 passing through the corners $x(0,0)$ and $x(M,0)$.

In order for a grid point $x(p,q)$ to be an acceptable continuation of a path through some previous point $x(p-1,q')$, it must satisfy two continuity conditions:

a) $|q-q'| \leq 1$; this condition restricts the path to follow non-negative time steps along the time coordinates of the patterns, and

b) $|q-q''| \leq 1$, where $x(p-2,q'')$ is the selected predecessor of the point $x(p-1,q')$; this condition restricts any one time frame to participation in at most two local comparisons.

With the aid of these constraints, each point in the search is restricted to at most three possible predecessors. To establish the minimal distance subpath $D_T(p,q)$ leading back to the origin from the point $x(p,q)$, the cumulative distance leading to that point through each possible predecessor $x(p-1,q')$ is minimized. Thus

$$D_T(p,q) = \min_{q'} \{D_T(p-1,q') = D[A(p-q),B(p+q)]\ V(q-q')\} \quad . \qquad (9)$$

V is a penalty function introduced to keep the alignment path close to the diagonal unless a significant distance reduction is obtained by following a different path. By setting V to 1.5 for $|q-q'| = 1$ and 1.0 otherwise, unproductive searches far from the diagonal are avoided. Since all paths terminate at x(M,0), the total distance of the minimum distance path and therefore the distance between A and B is given by $D_T(M,0)$.

The minimal distance subpath passes through the points $\{x(p,\hat{q}),\ 1 \leq p \leq M\}$. These points allow the identification of pairs of frames $A(p-\hat{q})$ and $B(p+\hat{q})$ that contributed to the minimal distance result. A new template C(p), p = 1,2,...,M, can then be generated by appropriately averaging the frames $A(p-\hat{q})$ and $B(p+\hat{q})$, p = 1,2,...,M.

The one exception to template generation by weighted averaging occurs with the LPC. If two LPC vectors are averaged, stability of the resultant vector is not guaranteed. Therefore, LPC templates were generated in the space of LP-derived reflection coefficients. Since the reflection coefficients are bounded in magnitude by one, stability requirements are satisfied and the symmetric dynamic warping algorithm could be used without modification. Alternately, the templates could be derived in the space of LP-derived autocorrelation coefficients, since stability is guaranteed from the result that a stable autocorrelation matrix is positive definite, and a linear combination of positive definite matrices is positive definite and hence stable.

3.4 Effects of Order In Generating a Template

As discussed above, the incremental addition of individual tokens to a previously formed template results in a final template whose values depend on the order of the tokens.

In a preliminary experiment utilizing the same data base (Davis, 1979), ten sets of reference templates based on six MFCC were generated. Each set of templates used the reference tokens in random order. Independent test data were then matched with each set of templates on a per speaker basis. The average recognition scores and standard deviations were 94.76 $\pm$ 0.53% and 90.53 $\pm$ 0.48% for each speaker respectively. Thus, random ordering of tokens for template generation did not change the results. At a 0.01 significance level, none of the rates for either speaker was significantly different from the respective mean. Thirty-two of the 52 different CVC word types were never misidentified. Errors were generally confined to the same tokens of a word regardless of the template, and the most confusions were among test-reference pairs such as wake-bait, book-hood and burn-herd.

The consistent rates among template sets indicated that the templates for any given word were relatively similar. To visualize such relationships, all of the pairwise distances for eight templates and four test tokens of keep were measured and fitted to an X-Y plane. The eight templates were arbitrarily chosen from among the 24 possible templates for four reference tokens from DZ1, and the four test tokens were obtained from DZ2. The fitting procedure

was based on iterating (x,y) coordinates for test each point (template or token) until the mean-square error in distances among the points was minimized. The coordinate plane is shown in Figure 5. Regardless of ordering, the templates are close to each other and relatively far from the test tokens, thus illustrating the robustness of the technique for template generation.

4. Recognition

For each parametric representation (MFCC, LFCC, LPCC, LPC and RC), the following test procedure was used (Davis & Mermelstein, 1978). Each segmented token from sets DZ1, DZ2, LL1 and LL2 was analyzed and a matrix of coefficients (columns corresponding to coefficient number and rows corresponding to time frame) was stored (Figure 6). Each set was used in turn as test and reference data. In the case of reference data, templates were formed on a per speaker per session basis, using all tokens of each word (generally three to five in number) recorded in the session. Two types of testing were used: closed tests, where test and reference data were from the same session, e.g., reference DZ1 vs. test DZ1, and open tests, where test and reference data were from different sessions, e.g., reference DZ1 vs. test DZ2 (Figure 7). For each test word, a warping was performed with each of the 52 reference templates, and the word was identified with the least distant template (maximum similarity). In a practical situation, alternative methods, such as vowel preselection and thresholding for early rejection, could be applied to reduce the computations and the number of comparisons. In this experiment, however, the emphasis was on methodology rather than efficiency.

The results are listed in Table 1 and displayed in Figure 8 for open tests with 10 coefficients and 6.4 ms frames. Regardless of the frame separation, type of testing or speaker, these data indicate superior performance of the MFCC when compared with the other parametric representations. In fact, the performance of six MFCC was also better than any other ten coefficient set. In all cases, the 6.4 ms frame separation produced better performance. As previously stated, the window size was 25.6 ms, and using half the window size produced degraded results. Finally, speaker DZ, a male with exceptionally low fundamental frequency, was better recognized than speaker LL, a male with somewhat higher fundamental frequency. Speaker dependent differences, however, require further systematic investigation.

Most confusions arose between pairs of words that were phonetically very similar. For example, of the eight misrecognitions using the MFCC parameters for speaker DZ, two were between "bar" and "mar," two were between "pool" and "fool," one each between "keep" and "heat," "bait" and "wake," "hook" and "rig," and "hood" and "cause." Note that by not using the average spectrum energy (the zeroth cepstrum coefficient) in these comparisons, the overall energy between time aligned spectral frames has been equalized. Inclusion of the variation of overall energy with time might possibly assist discrimination between such highly confusable word pairs.

Table 1

Recognition Rates Resulting from Use of Various Acoustic Representations

Acoustic Representation	Number of Coefficients	Distance Metric	Frame Separation (ms)	Speaker	Open Test %	Closed Test %
mel-frequency cepstrum	10	Euclidean	6.4	DZ	96.5	99.4
				LL	95.0	99.1
			12.8	DZ	95.6	99.4
				LL	93.8	97.9
mel-frequency cepstrum	6	Euclidean	6.4	DZ	96.5	99.4
				LL	92.0	97.6
			12.8	DZ	95.0	98.8
				LL	90.2	97.6
linear-frequency cepstrum	10	Euclidean	6.4	DZ	94.7	99.1
				LL	87.6	98.2
			12.8	DZ	93.2	98.8
				LL	84.9	97.3
linear-prediction cepstrum	10	Euclidean	6.4	DZ	92.6	99.1
				LL	87.3	98.2
			12.8	DZ	91.7	98.2
				LL	86.4	96.7
linear-prediction spectrum	10	Itakura	6.4	DZ	85.2	97.9
				LL	84.3	95.2
reflection coefficients	10	Euclidean	6.4	DZ	83.1	97.1
				LL	77.5	97.0
			12.8	DZ	80.5	97.6
				LL	74.6	96.2

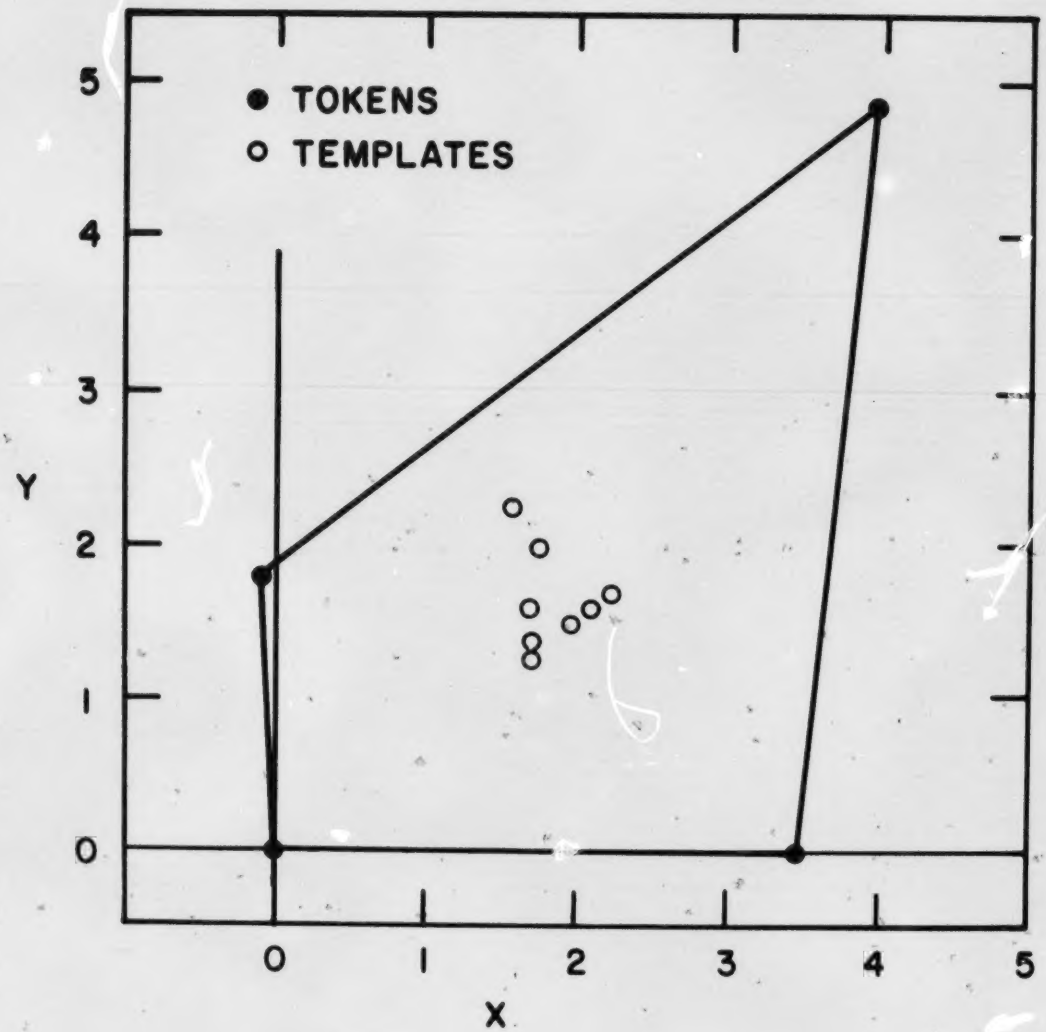


Figure 5: X-Y coordinate plane for keep.

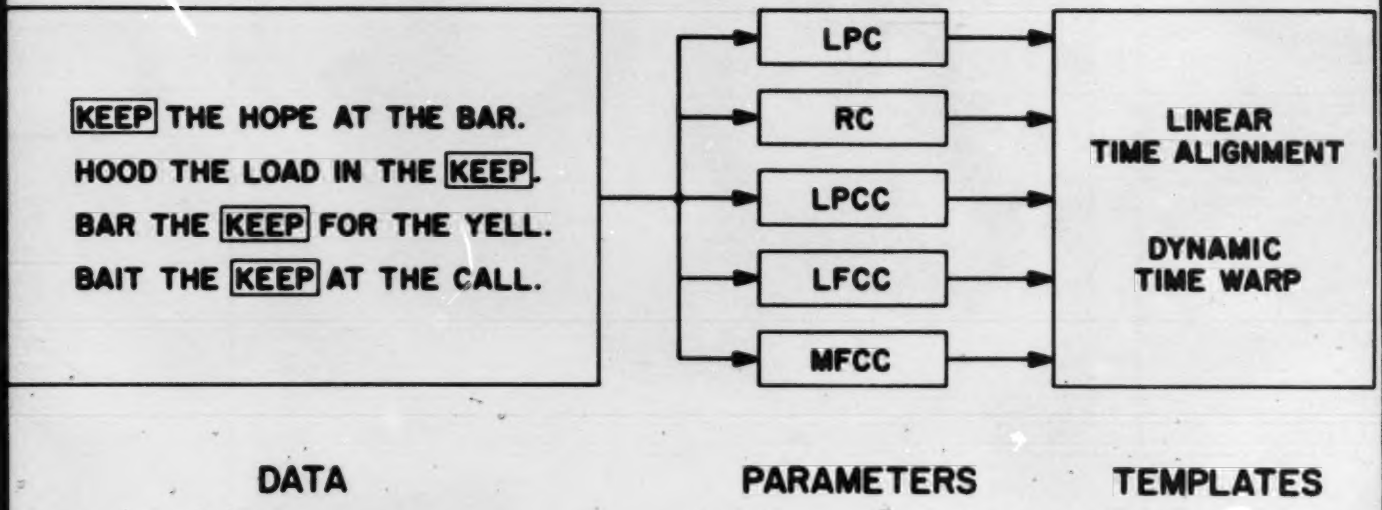


Figure 6: Selection of monosyllabic words for template generation.

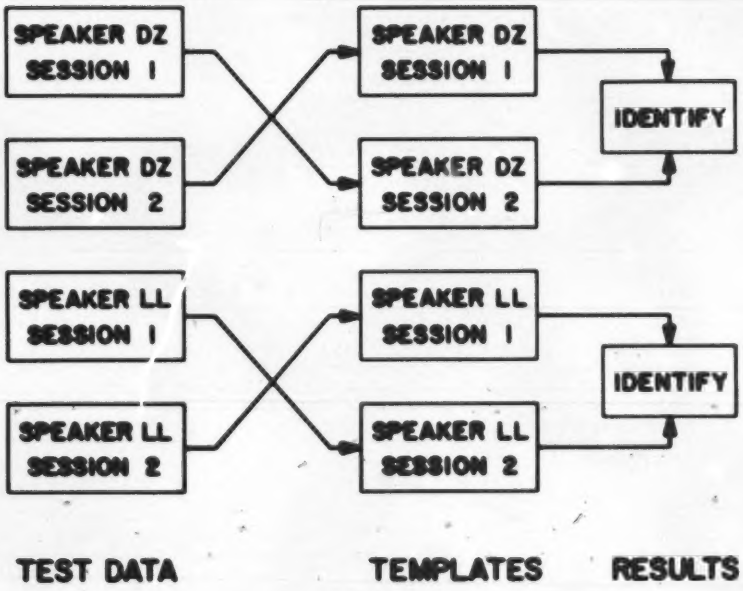
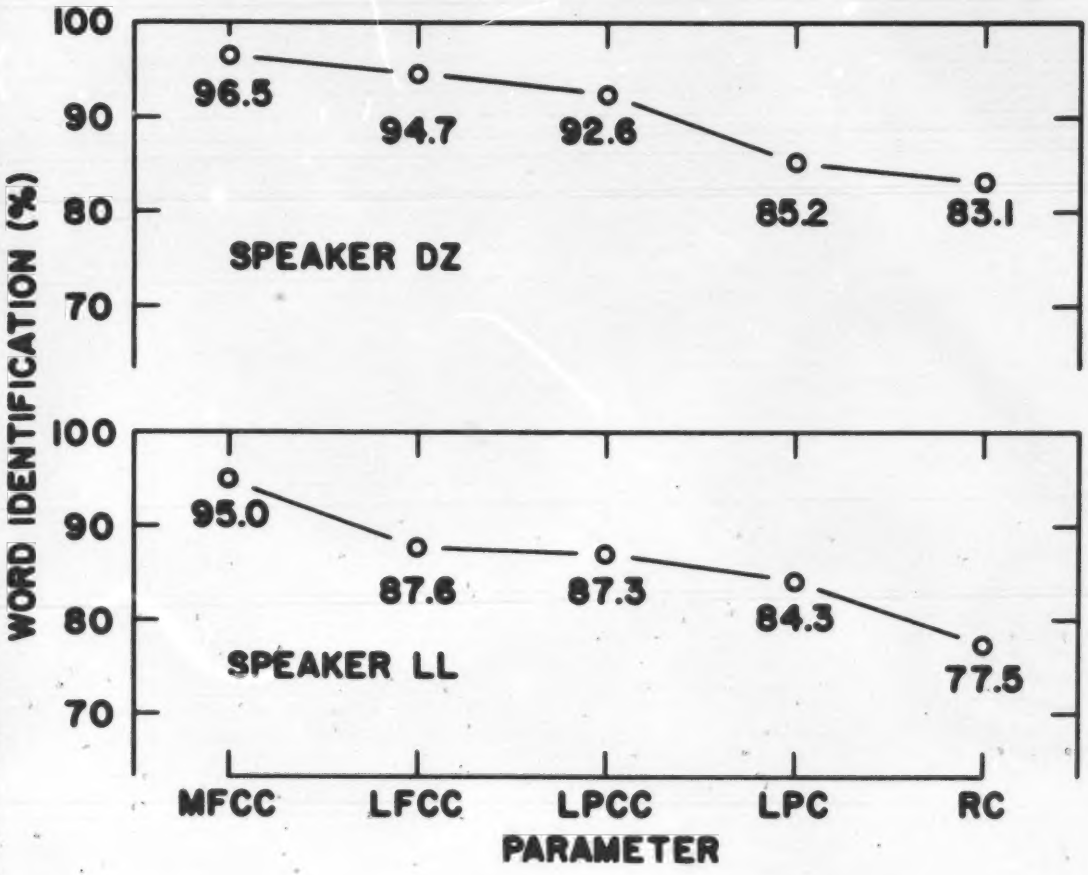


Figure 7: Two-way speaker-dependent identification tests.

Figure 8: Performance of parametric representations for recognition.



5. Conclusions

The similarity in rank order of the recogniton rates by representation for each of the two speakers suggests that the performance differences among the various acoustic representations are significant. These differences lead to the following specific conclusions:

1) Parameters derived from the short-term Fourier spectrum (MFCC, LFCC) of the acoustic signal preserve information that parameters from the LPC spectrum (LPCC, LPC, RC) omit. Both spectral representations are considered adequate for vowels. However, it is the confusions between the consonants that are most frequent. The differences found may be due to the insufficiently accurate representation of the consonantal spectra by the linear prediction technique.

2) The mel-frequency cepstra possess a significant advantage over the linear-frequency cepstra--specifically, MFCC allow better suppression of insignificant spectral variation in the higher frequency bands.

3) The cepstrum parameters (MFCC, LFCC and LPCC), which correspond to various frequency smoothed representations of the log-magnitude spectrum, succeed better than the LPC and RC in capturing the significant acoustic information. A Euclidean distance metric defined on the cepstrum parameters apparently allows a better separation of phonetically distinct spectra. Since there is a unique transformation between a set of LPCC and the corresponding LPC and RC, these representations can be said to contain equivalent information. However, this transformation is nonlinear. Representing the acoustic information in the hyperspace of cepstrum parameters favors the use of a particularly simple distance metric.

4) Defining the metric on the basis of the Itakura distance is less effective than defining it on the basis of cepstrum distance. The point of optimality is the same, i.e., equality between cepstra implies zero difference in prediction residual energy. However, the Itakura distance is less successful in indicating the phonetic significance of the difference between a pair of spectra than the cepstrum distance.

5) The mel-frequency cepstrum coefficients form a particularly compact representation. Six coefficients succeed in capturing most of the relevant information. The importance of the higher cepstrum coefficients appears to depend on the speaker. Further data are required from additional speakers before firm conclusions can be reached on the optimal number of coefficients.

The results are limited by the restrictions on the speech data examined. In particular, consonant clusters, multisyllabic words and unstressed monosyllabic words have not been studied. Expansion of the data base along any one of these directions introduces additional representation problems. It is not obvious that the best representation for stressed words is also best for the much more elastic unstressed words. These questions are left for future studies.

It should be emphasized that the comparative ranking of the representations can be influenced by the choice of both the local and the integrated distance metrics. A Euclidean distance function is one of the simplest to implement. However, taking into account the probability distributions of the individual parameters should result in improved performance. Estimating these distributions requires considerable data. Yet, even if only a few parameters of these distributions are known, for example, the variance of the cepstrum coefficients, better local distance metrics could be designed. Despite the high recognition rates achieved so far, there is reason to believe that even better performance can be attained in the future.

The design of the mel-frequency cepstrum representation was motivated by perceptual factors. Evidently, an ability to capture the perceptually relevant information is an important advantage. The design of an improved distance metric may result from more accurate modeling of perceptual behavior. In particular, where a constant difference between spectra persists for a number of consecutive time frames, the contribution of that difference in the current distance computation is proportional to the duration of that difference. With the possible exception of very short durations, no perceptual justification exists for this property (Feldtkeller & Zwicker, 1956). Nevertheless, the distance function must in some fashion combine different information from all the time frames constituting the signals compared. Further optimization of the integrated distance function represents an important challenge.

For each representation a small but significant gain in recognition is achieved by decreasing the frame spacing from 12.8 ms to 6.4 ms samples. The average difference in the recognition rates is 1.7%. However, the computational complexity for any dynamic programming comparison varies as the square of the average number of frames constituting a word. Thus a significant computational penalty accompanies any increase in the frame rate. In contrast, the computations grow only linearly with the number of cepstrum coefficients. Since the recognition rates for six cepstrum coefficients and 6.4 ms frame spacing is quite comparable to the rate for ten coefficients and 12.8 ms frame spacing, increasing the number of coefficients and maintaining a somewhat coarser time resolution is computationally more advantageous than using fewer coefficients more frequently.

The principal conclusion of the study is that perceptually based word templates are effective in capturing the acoustic information required to recognize these words in continuous speech. Due to the various limitations of this study, a conclusion that such high recognition rates are attainable with a complete automatic system operating in a practical environment is not warranted at this time. However, the results do encourage a continuing effort to optimize the performance of speech recognition systems by critical evaluation of each of the constituent components.

REFERENCES

Bridle, J. S., & Brown, M. D. An experimental automatic word recognition system. JSRU Report (Joint Speech Research Unit, Ruislip, England), 1974, No. 1003.

Davis, S. B. Order dependence in templates for monosyllabic word identifica-

tion. Conference Record, 1979 International Conference on Acoustics, Speech and Signal Processing, Washington, 1979, 570-573.

Davis, S. B., & Mermelstein, P. Evaluation of acoustic parameters for monosyllabic word identification. Journal of the Acoustical Society of America, 1978, 64, Suppl. 1, S180. (Abstract)

Fant, C. G. M. Acoustic description and classification of phonetic units. Ericsson Technics, 1959, 1. Also in G. Fant, Speech sounds and features, MIT Press, 32-83, 1973.

Feldtkeller, R., & Zwicker, E. Das Ohr als Nachrichtenempfanger. Stuttgart: S. Hirzel, 1956.

Fujimura, O. The syllable as a unit of speech recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, ASSP-23, 82-87.

Gray, A. H. Jr., & Markel, J. D. Distance measures for speech processing. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, ASSP-24, 380-391.

Itakura, F. Minimum prediction residual principle applied to speech recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, ASSP-23, 67-72.

Markel, J. D., & Gray, A. H. Jr. Linear prediction of speech. New York: Springer-Verlag, 1976.

Mermelstein, P. Automatic segmentation of speech into syllabic units. Journal of the Acoustical Society of America, 1975, 58, 880-883. (a)

Mermelstein, P. A phonetic-context controlled strategy for segmentation and phonetic labelling of speech. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, ASSP-23, 79-82. (b)

Mermelstein, P. Distance measures for speech recognition, psychological and instrumental. In C. H. Chen (Ed.), Pattern recognition and artificial intelligence. New York: Academic Press, 1976, 374-388.

Mermelstein, P. Recognition of monosyllabic words in continuous sentences using composite word templates. Conference Record, 1978 International Conference on Acoustics, Speech and Signal Processing, Tulsa, 1978, 708-711.

Pfeifer, L. L. Interactive laboratory system users guide. Santa Barbara: Signal Technology, Inc., 1977.

Pols, L. C. W. Spectral analysis and identification of Dutch vowels in monosyllabic words. Unpublished doctoral dissertation, Free University, Amsterdam, 1977.

Rabiner, L. R. On creating reference templates for speaker independent recognition of isolated words. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, ASSP-26, 34-42.

Rabiner, L. R., Rosenberg, A. E., & Levinson, S. E. Considerations in dynamic time warping algorithms for discrete word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, ASSP-26, 575-586.

Sakoe, H., & Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, ASSP-26, 43-49.

Schroeder, M. R. Recognition of complex acoustic signals. Life Sciences Research Report, T. H. Bullock ed., 1977, 55, 323-328.

Velichko, V. M., & Zagoruyko, N. G. Automatic recognition of 200 words. International Journal of Man-Machine Studies, 1970, 2, 223-234.

Wakita, H. Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms. IEEE Transactions on Acoustics, Speech and Signal Processing, 1973, AU-21, 417-427.

White, G. M., & Neely, R. B. Speech recognition experiments with linear prediction, bandpass filtering and dynamic programming. IEEE Transactions on Acoustics, Speech and Signal Processing, 1976, ASSP-24, 173-188.

FOOTNOTE

[1]Fant (1973) compares Beranek's mel-frequency scale, Koenig's scale and Fant's approximation to the mel-frequency scale. Since the differences between these scales are not significant here, the mel-frequency scale should be understood as a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz.

APPENDIX A

Sentences used for word recognition.

1. Keep the hope at the bar.
2. Dig this rock in the heat.
3. Wake the herd at the head.
4. Check the lock on the seal.
5. Bang this bar on the head.
6. Call a mess in the case.
7. Cut the coat for a mop.
8. Foot the work in the mess.
9. Boot the back of the book.
10. Burn your check in the jar.
11. Mop the room on the watch.
12. Load the tar for the bait.
13. Tar this rig in a rush.
14. Fear a hood on the ship.
15. Rig a bait for the work.
16. Nail that book to the rock.
17. Yell this call for the wake.
18. Gang the bait on the coat.
19. Walk the watch in the hope.
20. Buff one book for the walk.
21. Hook the mop on the lock.
22. Pool the case for the man.
23. Hurl his bar in the muck.
24. Bomb the head at the wake.
25. Pose this seal for the gang.
26. Mar the watch on the hood.
27. Heat the foot of the fool.
28. Kill the herd for the load.
29. Case your ship for the cause.
30. Head the rush for the burn.
31. Back the pool for the check.
32. Watch that hook with the nail.
33. Rush the buff at the foot.
34. Hood the load for the keep.
35. Room one seal in the pool.
36. Herd the fool with a yell.
37. Rock the mop with a hurl.
38. Coat the cut with the tar.
39. Jar the bomb with a bang.
40. Seal the dig in a fear.
41. Ship the nail in a boot.
42. Bait the keep with a call.
43. Mess his work in the room.
44. Man the cut at the kill.
45. Cause a mar on the back.
46. Muck the gang on the walk.
47. Book the fool on the rig.
48. Fool the man on the rock.
49. Work the hurl at the dig.
50. Lock your man in a pose.
51. Hope this call for the heat.
52. Bar the keep for the yell.
53. Put a bang in the bomb.
54. Set a pose in the muck.
55. Pose a jar on the buff.
56. Kill the fear in the cause.
57. Mar the burn on the head.

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

Borden, G., & Harris, K. S. Speech science primer. Baltimore, Md.: Williams & Wilkins, 1980.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop-consonant manner. Perception & Psychophysics, in press.

Healy, A. F. Proofreading errors on the word the: New evidence on reading units. Journal of Experimental Psychology: Human Perception and Performance, 1980, 6, 45-57.

Kelso, J. A. S. Remarks on preparatory processes. In E. Donchin (Ed.), Proceedings of conference on event-related brain potentials in the study of cognitive function. Hillsdale, N.J.: Lawrence Erlbaum, in press.

Repp, B. H. Stimulus dominance in fused dichotic syllables: Trouble for the category goodness hypothesis. Journal of the Acoustical Society of America, 1980, 67, 288-305.

Wexler, B. E., Halwes, T., & Heninger, G. R. Use of a statistical significance criterion in drawing inferences about cerebral dominance for language function from dichotic listening data. Brain and Language, in press.

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A013325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	AD A041460	ED-141-864
SR-50	April - June 1977	AD A044820	ED-144-138
SR-51/52	July - December 1977	AD A049215	ED-147-892
SR-53	January - March 1978	AD A055853	ED-155-760
SR-54	April - June 1978	AD A067070	ED-161-096
SR-55/56	July - December 1978	AD A065575	ED-166-75
SR-57	January - March 1979	AD A083179	ED-170-8
SR-58	April - June 1979	AD A077663	ED-178-9
SR-59/60	July - December 1979	AD A082034	ED-181

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

**DDC and/or ERIC order numbers not yet assigned.